

Document made available under the Patent Cooperation Treaty (PCT)

International application number: PCT/US2005/010912

International filing date: 31 March 2005 (31.03.2005)

Document type: Certified copy of priority document

Document details: Country/Office: US
Number: 60/559,040
Filing date: 01 April 2004 (01.04.2004)

Date of receipt at the International Bureau: 12 August 2005 (12.08.2005)

Remark: Priority document submitted or transmitted to the International Bureau in compliance with Rule 17.1(a) or (b)



World Intellectual Property Organization (WIPO) - Geneva, Switzerland
Organisation Mondiale de la Propriété Intellectuelle (OMPI) - Genève, Suisse

100291



THE UNITED STATES OF AMERICA

TO ALL TO WHOM THESE PRESENTS SHALL COME:

UNITED STATES DEPARTMENT OF COMMERCE

United States Patent and Trademark Office

August 02, 2005

THIS IS TO CERTIFY THAT ANNEXED HERETO IS A TRUE COPY FROM THE RECORDS OF THE UNITED STATES PATENT AND TRADEMARK OFFICE OF THOSE PAPERS OF THE BELOW IDENTIFIED PATENT APPLICATION THAT MET THE REQUIREMENTS TO BE GRANTED A FILING DATE.

APPLICATION NUMBER: 60/559,040

FILING DATE: April 01, 2004

RELATED PCT APPLICATION NUMBER: PCT/US05/10912



Certified by

Don W. Dudas

Under Secretary of Commerce
for Intellectual Property
and Director of the United States
Patent and Trademark Office

15750 U.S. PTO

PTO/SB/16 (08-03)

Approved for use through 7/31/2006. OMB 0651-0032
 U.S. Patent and Trademark Office: U.S. DEPARTMENT OF COMMERCE
 Under the Paperwork Reduction Act of 1995, no persons are required to respond to a collection of information unless it displays a valid OMB control number.

PROVISIONAL APPLICATION FOR PATENT COVER SHEET

This is a request for filing a PROVISIONAL APPLICATION FOR PATENT under 37 CFR 1.53(c).

Express Mail Label No. EL 961008081 US

22858 U.S. PTO
60/559040

INVENTOR(S)			
Given Name (first and middle [if any])	Family Name or Surname	Residence (City and either State or Foreign Country)	
Steven Andreas Stefan M.	MAH BRAUN KAMMERER	San Diego, California San Diego, California San Diego, California	
Additional inventors are being named on the <u>1</u> separately numbered sheets attached hereto			
TITLE OF THE INVENTION (500 characters max)			
METHODS FOR IDENTIFYING RISK OF OSTEOARTHRITIS AND TREATMENTS THEREOF			
Direct all correspondence to: CORRESPONDENCE ADDRESS			
<input checked="" type="checkbox"/> Customer Number:	25225		
OR			
<input type="checkbox"/> Firm or Individual Name			
Address			
City	State	Zip	
Country	Telephone	Fax	
ENCLOSED APPLICATION PARTS (check all that apply)			
<input checked="" type="checkbox"/> Specification Number of Pages	108	<input type="checkbox"/> CD(s), Number	
<input checked="" type="checkbox"/> Drawing(s) Number of Sheets	1	<input checked="" type="checkbox"/> Other	Return Receipt Postcard
<input checked="" type="checkbox"/> Application Data Sheet. See 37 CFR 1.76 (4 pages) (specify):			
METHOD OF PAYMENT OF FILING FEES FOR THIS PROVISIONAL APPLICATION FOR PATENT			
<input checked="" type="checkbox"/> Applicant claims small entity status. See 37 CFR 1.27.		FILING FEE AMOUNT (\$)	
<input type="checkbox"/> A check or money order is enclosed to cover the filing fees.		80.00	
<input checked="" type="checkbox"/> The Director is hereby authorized to charge filing fees or credit any overpayment to Deposit Account Number:		03-1952	
<input type="checkbox"/> Payment by credit card. Form PTO-2038 is attached.			
The invention was made by an agency of the United States Government or under a contract with an agency of the United States Government.			
<input checked="" type="checkbox"/> No <input type="checkbox"/> Yes, the name of the U.S. Government agency and the Government contract number are:			

(Page 1 of 2)

Respectfully submitted,

Date April 1, 2004

SIGNATURE

TYPED OR

PRINTED NAME

Bruce D. Grant

TELEPHONE

(858) 720-7962

REGISTRATION NO.

47,608

(if appropriate)

Docket Number:

524593009100

USE ONLY FOR FILING A PROVISIONAL APPLICATION FOR PATENT

I hereby certify that this correspondence is being deposited with the U.S. Postal Service as Express Mail, Airbill No. EL 961008081 US, in an envelope addressed to: Mail Stop Provisional Patent Application, Commissioner for Patents, P.O. Box 1450, Alexandria, VA 22313-1450, on the date shown below.

Dated: 4/1/04

Signature: Deborah Wykes (Deborah Wykes)

Additional Page

PTO/SB/16 (08-03)

Approved for use through 07/31/06. OMB 0651-0032

U.S. Patent and Trademark Office: U.S. DEPARTMENT OF COMMERCE

Under the Paperwork Reduction Act of 1995, no persons are required to respond to a collection of information unless it displays a valid OMB control number.

Docket Number 524593009100

INVENTOR(S)/APPLICANT(S)		
Given Name (first and middle (if any))	Family or Surname	Residence (City and either State or Foreign Country)
Matthew Roberts Rikard Henry Maria L.	NELSON RENELAND LANGDOWN	San Marcos, California San Diego, California San Diego, California

**METHODS FOR IDENTIFYING RISK OF OSTEOARTHRITIS AND
TREATMENTS THEREOF**

Field of the Invention

[0001] The invention relates to genetic methods for identifying risk of osteoarthritis and treatments that specifically target such diseases.

Background

[0002] Osteoarthritis (OA) is a chronic disease usually affecting weight-bearing synovial joints. There are approximately 20 million Americans affected by OA and it is the leading cause of disability in the United States. In addition to extensive human suffering, OA also accounts for nearly all knee replacements and more than half of all hip replacements in the United States. Despite its prevalence, OA is poorly understood and there are few treatments available besides anti-inflammatory drugs and joint replacement.

[0003] Most commonly affecting middle-aged and older people, OA can range from very mild to very severe. It affects hands and weight-bearing joints such as knees, hips, feet and the back. Knee OA can be as disabling as any cardiovascular disease except stroke.

[0004] OA is characterized by the breakdown of cartilage in joints. Cartilage in joints cushions the ends of bones, and cartilage breakdown causes bones to rub against each other, causing pain and loss of movement. Type II collagen is the main component of cartilage, comprising 15-25% of the wet weight, approximately half the dry weight, and representing 90-95% of the total collagen content in the tissue. It forms fibrils that endow cartilage with tensile strength (Mayne, R. Arthritis Rheum. 32:241-246 (1989)).

Summary

[0005] It has been discovered that certain polymorphic variations in human genomic DNA are associated with osteoarthritis. In particular, polymorphic variants in a locus containing a *TM7SF3* region in human genomic DNA have been associated with risk of osteoarthritis.

[0006] Thus, featured herein are methods for identifying a subject at risk of osteoarthritis and/or a risk of osteoarthritis in a subject, which comprise detecting the presence or absence of one or more polymorphic variations associated with osteoarthritis in or around the loci described herein in a human nucleic acid sample. In an embodiment, two or more polymorphic variations are detected in two or more regions of which one is the *TM7SF3* region. In certain embodiments, 3 or more, or 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19 or 20 or more polymorphic variants are detected.

[0007] Also featured are nucleic acids that include one or more polymorphic variations associated with occurrence of osteoarthritis, as well as polypeptides encoded by these nucleic acids. In addition, provided are methods for identifying candidate therapeutic molecules for treating osteoarthritis, as well as methods for treating osteoarthritis in a subject by identifying a subject at risk of osteoarthritis and treating the subject with a suitable prophylactic, treatment or therapeutic molecule.

[0008] Also provided are compositions comprising a cell from a subject having osteoarthritis or at risk of osteoarthritis and/or a *TM7SF3* nucleic acid, with a RNAi, siRNA, antisense DNA or RNA, or ribozyme nucleic acid designed from a *TM7SF3* nucleotide sequence. In an embodiment, the RNAi, siRNA, antisense DNA or RNA, or ribozyme nucleic acid is designed from a *TM7SF3* nucleotide sequence that includes one or more polymorphic variations associated with osteoarthritis, and in some instances, specifically interacts with such a nucleotide sequence. Further, provided are arrays of nucleic acids bound to a solid surface, in which one or more nucleic acid molecules of the array have a *TM7SF3* nucleotide sequence, or a fragment or substantially identical nucleic acid thereof, or a complementary nucleic acid of the foregoing. Featured also are compositions comprising a cell from a subject having osteoarthritis or at risk of osteoarthritis and/or a *TM7SF3* polypeptide, with an antibody that specifically binds to the polypeptide. In an embodiment, the antibody specifically binds to an epitope in the polypeptide that includes a non-synonymous amino acid modification associated with osteoarthritis (e.g., results in an amino acid substitution in the encoded polypeptide associated with osteoarthritis).

Brief Description of the Drawings

[0009] Figure 1 shows proximal SNPs in a *TM7SF3* region in genomic DNA. The position of each SNP in the chromosome is shown on the x-axis and the y-axis provides the negative logarithm of the p-value comparing the estimated allele to that of the control group. Also shown in the figure are exons and introns of the gene region in the approximate chromosomal positions.

Detailed Description

[0010] It has been discovered that a polymorphic variant in a locus containing a *TM7SF3* region is associated with occurrence of osteoarthritis in subjects. Thus, detecting genetic determinants associated with an increased risk of osteoarthritis occurrence can lead to early identification of a predisposition to osteoarthritis and early prescription of preventative measures. Also, associating a *TM7SF3* polymorphic variant with osteoarthritis has provided new targets for screening molecules useful in treatments of osteoarthritis.

Osteoarthritis and Sample Selection

[0011] Osteoarthritis (OA), or degenerative joint disease, is one of the oldest and most common types of arthritis. It is characterized by the breakdown of the joint's cartilage. Cartilage is the part of the joint that cushions the ends of bones, and its breakdown causes bones to rub against each other, causing pain and loss of movement. Type II collagen is the main component of cartilage, comprising 15-25% of the wet weight, approximately half the dry weight, and representing 90-95% of the total collagen content in the tissue. It forms fibrils that endow cartilage with tensile strength (Mayne, R. Arthritis Rheum. 32:241-246 (1989)).

[0012] Most commonly affecting middle-aged and older people, OA can range from very mild to very severe. It affects hands and weight-bearing joints such as knees, hips, feet and the back. Knee OA can be as disabling as any cardiovascular disease except stroke.

[0013] Osteoarthritis affects an estimated 20.7 million Americans, mostly after age 45, with women more commonly affected than men. Physicians make a diagnosis of OA based on a physical exam and history of symptoms. X-rays are used to confirm diagnosis. Most people over 60 reflect the disease on X-ray, and about one-third have actual symptoms.

[0014] There are many factors that can cause OA. Obesity may lead to osteoarthritis of the knees. In addition, people with joint injuries due to sports, work-related activity or accidents may be at increased risk of developing OA.

[0015] Genetics has a role in the development of OA. Some people may be born with defective cartilage or with slight defects in the way that joints fit together. As a person ages, these defects may cause early cartilage breakdown in the joint or the inability to repair damaged or deteriorated cartilage in the joint.

[0016] Inclusion or exclusion of samples for an osteoarthritis pool may be based upon the following criteria: ethnicity (e.g., samples derived from an individual characterized as Caucasian); parental ethnicity (e.g., samples derived from an individual of British paternal and maternal descent); relevant phenotype information for the individual (e.g., case samples derived from individuals diagnosed with specific knee osteoarthritis (OA) and were recruited from an OA knee replacement clinic). Control samples may be selected based on relevant phenotype information for the individual (e.g., derived from individuals free of OA at several sites (knee, hand, hip etc)); and no family history of OA and/or rheumatoid arthritis. Additional phenotype information collected for both cases and controls may include age of the individual, gender, family history of OA, diagnosis with osteoarthritis (joint location of OA, date of primary diagnosis, age of individual as of primary diagnosis), knee history (current symptoms, any major knee injury, meniscectomy, knee replacement surgery, age of surgery), HRT history, osteoporosis diagnosis.

[0017] Based in part upon selection criteria set forth above, individuals having osteoarthritis can be selected for genetic studies. Also, individuals having no history of osteoarthritis often are selected for genetic studies, as described hereafter.

Polymorphic Variants Associated with Osteoarthritis

[0018] A genetic analysis provided herein linked osteoarthritis with polymorphic variant nucleic acid sequences in the human genome. As used herein, the term “polymorphic site” refers to a region in a nucleic acid at which two or more alternative nucleotide sequences are observed in a significant number of nucleic acid samples from a population of individuals. A polymorphic site may be a nucleotide sequence of two or more nucleotides, an inserted nucleotide or nucleotide sequence, a deleted nucleotide or nucleotide sequence, or a microsatellite, for example. A polymorphic site that is two or more nucleotides in length may be 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15 or more, 20 or more, 30 or more, 50 or more, 75 or more, 100 or more, 500 or more, or about 1000 nucleotides in length, where all or some of the nucleotide sequences differ within the region. A polymorphic site is often one nucleotide in length, which is referred to herein as a “single nucleotide polymorphism” or a “SNP.”

[0019] Where there are two, three, or four alternative nucleotide sequences at a polymorphic site, each nucleotide sequence is referred to as a “polymorphic variant” or “nucleic acid variant.” Where two polymorphic variants exist, for example, the polymorphic variant represented in a minority of samples from a population is sometimes referred to as a “minor allele” and the polymorphic variant that is more prevalently represented is sometimes referred to as a “major allele.” Many organisms possess a copy of each chromosome (*e.g.*, humans), and those individuals who possess two major alleles or two minor alleles are often referred to as being “homozygous” with respect to the polymorphism, and those individuals who possess one major allele and one minor allele are normally referred to as being “heterozygous” with respect to the polymorphism. Individuals who are homozygous with respect to one allele are sometimes predisposed to a different phenotype as compared to individuals who are heterozygous or homozygous with respect to another allele.

[0020] In genetic analysis that associate polymorphic variants with osteoarthritis, samples from individuals having osteoarthritis and individuals not having osteoarthritis often are allelotyped and/or genotyped. The term “allelotype” as used herein refers to a process for determining the allele frequency for a polymorphic variant in pooled DNA samples from cases and controls. By pooling DNA from each group, an allele frequency for each SNP in each group is calculated. These allele frequencies are then compared to one another. The term “genotyped” as used herein refers to a process for determining a genotype of one or more individuals, where a “genotype” is a representation of one or more polymorphic variants in a population.

[0021] A genotype or polymorphic variant may be expressed in terms of a “haplotype,” which as used herein refers to two or more polymorphic variants occurring within genomic DNA in a group of individuals within a population. For example, two SNPs may exist within a gene where each SNP position includes a cytosine variation and an adenine variation. Certain individuals in a population may carry one allele (heterozygous) or two alleles (homozygous) having the gene with a cytosine at each SNP position. As the two cytosines corresponding to each SNP in the gene travel together on one or both alleles in these individuals, the individuals can be characterized as having a cytosine/cytosine haplotype with respect to the two SNPs in the gene.

[0022] As used herein, the term “phenotype” refers to a trait which can be compared between individuals, such as presence or absence of a condition, a visually observable difference in appearance between individuals, metabolic variations, physiological variations, variations in the function of biological molecules, and the like. An example of a phenotype is occurrence of osteoarthritis.

[0023] Researchers sometimes report a polymorphic variant in a database without determining whether the variant is represented in a significant fraction of a population. Because a subset of these reported polymorphic variants are not represented in a statistically significant portion of the population, some of them are sequencing errors and/or not biologically relevant. Thus, it is often not known whether a reported polymorphic variant is statistically significant or biologically relevant until the presence of the variant is detected in a population of individuals and the frequency of the variant is determined. Methods for detecting a polymorphic variant in a population are described herein, specifically in Example 2. A polymorphic variant is statistically significant and often biologically relevant if it is represented in 5% or more of a population, sometimes 10% or more, 15% or more, or 20% or more of a population, and often 25% or more, 30% or more, 35% or more, 40% or more, 45% or more, or 50% or more of a population.

[0024] A polymorphic variant may be detected on either or both strands of a double-stranded nucleic acid. Also, a polymorphic variant may be located within an intron or exon of a gene or within a portion of a regulatory region such as a promoter, a 5' untranslated region (UTR), a 3' UTR, and in DNA (e.g., genomic DNA (gDNA) and complementary DNA (cDNA)), RNA (e.g., mRNA, tRNA, and rRNA), or a polypeptide. Polymorphic variations may or may not result in detectable differences in gene expression, polypeptide structure, or polypeptide function.

[0025] It was determined that polymorphic variations associated with an increased risk of osteoarthritis existed in a *TM7SF3* region in SEQ ID NO: 1. In certain embodiments, a polymorphic variant at position rs1484086 in the human genome was associated with an increased risk of osteoarthritis, and in a specific embodiment, a thymine at position rs1484086 was associated with an increased risk of osteoarthritis.

[0026] Polymorphic variants in and around the *TM7SF3* locus were tested for association with osteoarthritis. These include polymorphic variants at positions in SEQ ID NO: 1 selected from the group

consisting of 230, 231, 5330, 6334, 11372, 11456, 11501, 13393, 16666, 17596, 19710, 19800, 20297, 20967, 32514, 33159, 37600, 41259, 41329, 50060, 53292, 53393, 56417, 56435, 58847, 59595, 59661, 60355, 60407, 62357, 68230, 68516, 69055, 72603, 73928, 85897 and 91554. Polymorphic variants at the following positions in SEQ ID NO: 1 in particular were associated with an increased risk of osteoarthritis: 56435, 59595, 53292, 33159 and 41329, with specific embodiments directed to positions 56435 and/or 59595. In particular, the following polymorphic variants in SEQ ID NO: 1 were associated with risk of osteoarthritis: a thymine thymine repeat at position 56435, a thymine at position 59595, a cytosine at position 53292, a guanine at position 33159 and a thymine at position 41329.

[0027] Based in part upon analyses summarized in Figure 1, a region with significant association has been identified in a locus associated with osteoarthritis. Any polymorphic variants associated with osteoarthritis in a region of significant association can be utilized for embodiments described herein. For example, polymorphic variants in a region approximately 14,000 nucleotides in length spanning chromosome positions 27052000 to 27066000 in a *TM7SF3* locus have significant association (chromosome positions are within NCBI's Genome build 34).

Additional Polymorphic Variants Associated with Osteoarthritis

[0028] Also provided is a method for identifying polymorphic variants proximal to an incident, founder polymorphic variant associated with osteoarthritis. Thus, featured herein are methods for identifying a polymorphic variation associated with osteoarthritis that is proximal to an incident polymorphic variation associated with osteoarthritis, which comprises identifying a polymorphic variant proximal to the incident polymorphic variant associated with osteoarthritis, where the incident polymorphic variant is in a *TM7SF3* nucleotide sequence. The nucleotide sequence often comprises a polynucleotide sequence selected from the group consisting of (a) a polynucleotide sequence of SEQ ID NO: 1-2; (b) a polynucleotide sequence that encodes a polypeptide having an amino acid sequence encoded by a polynucleotide sequence of SEQ ID NO: 1-2; and (c) a polynucleotide sequence that encodes a polypeptide having an amino acid sequence that is 90% or more identical to an amino acid sequence encoded by a nucleotide sequence of SEQ ID NO: 1-2 or a polynucleotide sequence 90% or more identical to the polynucleotide sequence of SEQ ID NO: 1-2. The presence or absence of an association of the proximal polymorphic variant with osteoarthritis then is determined using a known association method, such as a method described in the Examples hereafter. In an embodiment, the incident polymorphic variant is a polymorphic variant associated with osteoarthritis described herein. In another embodiment, the proximal polymorphic variant identified sometimes is a publicly disclosed polymorphic variant, which for example, sometimes is published in a publicly available database. In other embodiments, the polymorphic variant identified is not publicly disclosed and is discovered using a known method, including, but not limited to, sequencing a region surrounding the incident polymorphic

variant in a group of nucleic samples. Thus, multiple polymorphic variants proximal to an incident polymorphic variant are associated with osteoarthritis using this method.

[0029] The proximal polymorphic variant often is identified in a region surrounding the incident polymorphic variant. In certain embodiments, this surrounding region is about 50 kb flanking the first polymorphic variant (e.g. about 50 kb 5' of the first polymorphic variant and about 50 kb 3' of the first polymorphic variant), and the region sometimes is composed of shorter flanking sequences, such as flanking sequences of about 40 kb, about 30 kb, about 25 kb, about 20 kb, about 15 kb, about 10 kb, about 7 kb, about 5 kb, or about 2 kb 5' and 3' of the incident polymorphic variant. In other embodiments, the region is composed of longer flanking sequences, such as flanking sequences of about 55 kb, about 60 kb, about 65 kb, about 70 kb, about 75 kb, about 80 kb, about 85 kb, about 90 kb, about 95 kb, or about 100 kb 5' and 3' of the incident polymorphic variant.

[0030] In certain embodiments, polymorphic variants associated with osteoarthritis are identified iteratively. For example, a first proximal polymorphic variant is associated with osteoarthritis using the methods described above and then another polymorphic variant proximal to the first proximal polymorphic variant is identified (e.g., publicly disclosed or discovered) and the presence or absence of an association of one or more other polymorphic variants proximal to the first proximal polymorphic variant with osteoarthritis is determined.

[0031] The methods described herein are useful for identifying or discovering additional polymorphic variants that may be used to further characterize a gene, region or loci associated with a condition, a disease (e.g., osteoarthritis), or a disorder. For example, allelotyping or genotyping data from the additional polymorphic variants may be used to identify a functional mutation or a region of linkage disequilibrium. In certain embodiments, polymorphic variants identified or discovered within a region comprising the first polymorphic variant associated with osteoarthritis are genotyped using the genetic methods and sample selection techniques described herein, and it can be determined whether those polymorphic variants are in linkage disequilibrium with the first polymorphic variant. The size of the region in linkage disequilibrium with the first polymorphic variant also can be assessed using these genotyping methods. Thus, provided herein are methods for determining whether a polymorphic variant is in linkage disequilibrium with a first polymorphic variant associated with osteoarthritis, and such information can be used in prognosis/diagnosis methods described herein.

Isolated Nucleic Acids

[0032] Featured herein are isolated *TM7SF3* nucleic acid variants depicted in SEQ ID NO: 1-2, and substantially identical nucleic acids thereof. A nucleic acid variant may be represented on one or both strands in a double-stranded nucleic acid or on one chromosomal complement (heterozygous) or both chromosomal complements (homozygous).

[0033] As used herein, the term “nucleic acid” includes DNA molecules (*e.g.*, a complementary DNA (cDNA) and genomic DNA (gDNA)) and RNA molecules (*e.g.*, mRNA, rRNA, siRNA and tRNA) and analogs of DNA or RNA, for example, by use of nucleotide analogs. The nucleic acid molecule can be single-stranded and it is often double-stranded. The term “isolated or purified nucleic acid” refers to nucleic acids that are separated from other nucleic acids present in the natural source of the nucleic acid. For example, with regard to genomic DNA, the term “isolated” includes nucleic acids which are separated from the chromosome with which the genomic DNA is naturally associated. An “isolated” nucleic acid is often free of sequences which naturally flank the nucleic acid (*i.e.*, sequences located at the 5' and/or 3' ends of the nucleic acid) in the genomic DNA of the organism from which the nucleic acid is derived. For example, in various embodiments, the isolated nucleic acid molecule can contain less than about 5 kb, 4 kb, 3 kb, 2 kb, 1 kb, 0.5 kb or 0.1 kb of 5' and/or 3' nucleotide sequences which flank the nucleic acid molecule in genomic DNA of the cell from which the nucleic acid is derived. Moreover, an “isolated” nucleic acid molecule, such as a cDNA molecule, can be substantially free of other cellular material, or culture medium when produced by recombinant techniques, or substantially free of chemical precursors or other chemicals when chemically synthesized. As used herein, the term “gene” refers to a nucleotide sequence that encodes a polypeptide.

[0034] Also included herein are nucleic acid fragments. These fragments often have a nucleotide sequence identical to a nucleotide sequence of SEQ ID NO: 1-2, a nucleotide sequence substantially identical to a nucleotide sequence of SEQ ID NO: 1-2, or a nucleotide sequence that is complementary to the foregoing. The nucleic acid fragment may be identical, substantially identical or homologous to a nucleotide sequence in an exon or an intron in a nucleotide sequence of SEQ ID NO: 1-2, and may encode a domain or part of a domain of a polypeptide. Sometimes, the fragment will comprises one or more of the polymorphic variations described herein as being associated with osteoarthritis. The nucleic acid fragment is often 50, 100, or 200 or fewer base pairs in length, and is sometimes about 300, 400, 500, 600, 700, 800, 900, 1000, 1100, 1200, 1300, 1400, 1500, 2000, 3000, 4000, 5000, 10000, 15000, or 20000 base pairs in length. A nucleic acid fragment that is complementary to a nucleotide sequence identical or substantially identical to a nucleotide sequence in SEQ ID NO: 1-2 and hybridizes to such a nucleotide sequence under stringent conditions is often referred to as a “probe.” Nucleic acid fragments often include one or more polymorphic sites, or sometimes have an end that is adjacent to a polymorphic site as described hereafter.

[0035] An example of a nucleic acid fragment is an oligonucleotide. As used herein, the term “oligonucleotide” refers to a nucleic acid comprising about 8 to about 50 covalently linked nucleotides, often comprising from about 8 to about 35 nucleotides, and more often from about 10 to about 25 nucleotides. The backbone and nucleotides within an oligonucleotide may be the same as those of naturally occurring nucleic acids, or analogs or derivatives of naturally occurring nucleic acids, provided

that oligonucleotides having such analogs or derivatives retain the ability to hybridize specifically to a nucleic acid comprising a targeted polymorphism. Oligonucleotides described herein may be used as hybridization probes or as components of prognostic or diagnostic assays, for example, as described herein.

[0036] Oligonucleotides are typically synthesized using standard methods and equipment, such as the ABI™3900 High Throughput DNA Synthesizer and the EXPEDITE™ 8909 Nucleic Acid Synthesizer, both of which are available from Applied Biosystems (Foster City, CA). Analogs and derivatives are exemplified in U.S. Pat. Nos. 4,469,863; 5,536,821; 5,541,306; 5,637,683; 5,637,684; 5,700,922; 5,717,083; 5,719,262; 5,739,308; 5,773,601; 5,886,165; 5,929,226; 5,977,296; 6,140,482; WO 00/56746; WO 01/14398, and related publications. Methods for synthesizing oligonucleotides comprising such analogs or derivatives are disclosed, for example, in the patent publications cited above and in U.S. Pat. Nos. 5,614,622; 5,739,314; 5,955,599; 5,962,674; 6,117,992; in WO 00/75372; and in related publications.

[0037] Oligonucleotides may also be linked to a second moiety. The second moiety may be an additional nucleotide sequence such as a tail sequence (e.g., a polyadenosine tail), an adapter sequence (e.g., phage M13 universal tail sequence), and others. Alternatively, the second moiety may be a non-nucleotide moiety such as a moiety which facilitates linkage to a solid support or a label to facilitate detection of the oligonucleotide. Such labels include, without limitation, a radioactive label, a fluorescent label, a chemiluminescent label, a paramagnetic label, and the like. The second moiety may be attached to any position of the oligonucleotide, provided the oligonucleotide can hybridize to the nucleic acid comprising the polymorphism.

Uses for Nucleic Acid Sequence

[0038] Nucleic acid coding sequences may be used for diagnostic purposes for detection and control of polypeptide expression. Also, included herein are oligonucleotide sequences such as antisense RNA, small-interfering RNA (siRNA) and DNA molecules and ribozymes that function to inhibit translation of a polypeptide. Antisense techniques and RNA interference techniques are known in the art and are described herein.

[0039] Ribozymes are enzymatic RNA molecules capable of catalyzing the specific cleavage of RNA. The mechanism of ribozyme action involves sequence specific hybridization of the ribozyme molecule to complementary target RNA, followed by endonucleolytic cleavage. For example, hammerhead motif ribozyme molecules may be engineered that specifically and efficiently catalyze endonucleolytic cleavage of RNA sequences corresponding to or complementary to *TM7SF3* nucleotide sequences. Specific ribozyme cleavage sites within any potential RNA target are initially identified by scanning the target molecule for ribozyme cleavage sites which include the following sequences, GUA,

GUU and GUC. Once identified, short RNA sequences of between fifteen (15) and twenty (20) ribonucleotides corresponding to the region of the target gene containing the cleavage site may be evaluated for predicted structural features such as secondary structure that may render the oligonucleotide sequence unsuitable. The suitability of candidate targets may also be evaluated by testing their accessibility to hybridization with complementary oligonucleotides, using ribonuclease protection assays.

[0040] Antisense RNA and DNA molecules, siRNA and ribozymes may be prepared by any method known in the art for the synthesis of RNA molecules. These include techniques for chemically synthesizing oligodeoxyribonucleotides well known in the art such as solid phase phosphoramidite chemical synthesis. Alternatively, RNA molecules may be generated by *in vitro* and *in vivo* transcription of DNA sequences encoding the antisense RNA molecule. Such DNA sequences may be incorporated into a wide variety of vectors which incorporate suitable RNA polymerase promoters such as the T7 or SP6 polymerase promoters. Alternatively, antisense cDNA constructs that synthesize antisense RNA constitutively or inducibly, depending on the promoter used, can be introduced stably into cell lines.

[0041] DNA encoding a polypeptide also may have a number of uses for the diagnosis of diseases, including osteoarthritis, resulting from aberrant expression of a target gene described herein. For example, the nucleic acid sequence may be used in hybridization assays of biopsies or autopsies to diagnose abnormalities of expression or function (*e.g.*, Southern or Northern blot analysis, *in situ* hybridization assays).

[0042] In addition, the expression of a polypeptide during embryonic development may also be determined using nucleic acid encoding the polypeptide. As addressed, *infra*, production of functionally impaired polypeptide is the cause of various disease states, such as osteoarthritis. *In situ* hybridizations using polypeptide as a probe may be employed to predict problems related to osteoarthritis. Further, as indicated, *infra*, administration of human active polypeptide, recombinantly produced as described herein, may be used to treat disease states related to functionally impaired polypeptide. Alternatively, gene therapy approaches may be employed to remedy deficiencies of functional polypeptide or to replace or compete with dysfunctional polypeptide.

Expression Vectors, Host Cells, and Genetically Engineered Cells

[0043] Provided herein are nucleic acid vectors, often expression vectors, which contain a *TM7SF3* nucleotide sequence, or a substantially identical sequence thereof. As used herein, the term "vector" refers to a nucleic acid molecule capable of transporting another nucleic acid to which it has been linked and can include a plasmid, cosmid, or viral vector. The vector can be capable of autonomous replication or it can integrate into a host DNA. Viral vectors may include replication defective retroviruses, adenoviruses and adeno-associated viruses for example.

[0044] A vector can include a *TM7SF3* nucleotide sequence in a form suitable for expression of an encoded target polypeptide or target nucleic acid in a host cell. A "target polypeptide" is a polypeptide encoded by a *TM7SF3* nucleotide sequence, or a substantially identical nucleotide sequence thereof. The recombinant expression vector typically includes one or more regulatory sequences operatively linked to the nucleic acid sequence to be expressed. The term "regulatory sequence" includes promoters, enhancers and other expression control elements (e.g., polyadenylation signals). Regulatory sequences include those that direct constitutive expression of a nucleotide sequence, as well as tissue-specific regulatory and/or inducible sequences. The design of the expression vector can depend on such factors as the choice of the host cell to be transformed, the level of expression of polypeptide desired, and the like. Expression vectors can be introduced into host cells to produce target polypeptides, including fusion polypeptides.

[0045] Recombinant expression vectors can be designed for expression of target polypeptides in prokaryotic or eukaryotic cells. For example, target polypeptides can be expressed in *E. coli*, insect cells (e.g., using baculovirus expression vectors), yeast cells, or mammalian cells. Suitable host cells are discussed further in Goeddel, *Gene Expression Technology: Methods in Enzymology* 185, Academic Press, San Diego, CA (1990). Alternatively, the recombinant expression vector can be transcribed and translated *in vitro*, for example using T7 promoter regulatory sequences and T7 polymerase.

[0046] Expression of polypeptides in prokaryotes is most often carried out in *E. coli* with vectors containing constitutive or inducible promoters directing the expression of either fusion or non-fusion polypeptides. Fusion vectors add a number of amino acids to a polypeptide encoded therein, usually to the amino terminus of the recombinant polypeptide. Such fusion vectors typically serve three purposes: 1) to increase expression of recombinant polypeptide; 2) to increase the solubility of the recombinant polypeptide; and 3) to aid in the purification of the recombinant polypeptide by acting as a ligand in affinity purification. Often, a proteolytic cleavage site is introduced at the junction of the fusion moiety and the recombinant polypeptide to enable separation of the recombinant polypeptide from the fusion moiety subsequent to purification of the fusion polypeptide. Such enzymes, and their cognate recognition sequences, include Factor Xa, thrombin and enterokinase. Typical fusion expression vectors include pGEX (Pharmacia Biotech Inc; Smith & Johnson, *Gene* 67: 31-40 (1988)), pMAL (New England Biolabs, Beverly, MA) and pRIT5 (Pharmacia, Piscataway, NJ) which fuse glutathione S-transferase (GST), maltose E binding polypeptide, or polypeptide A, respectively, to the target recombinant polypeptide.

[0047] Purified fusion polypeptides can be used in screening assays and to generate antibodies specific for target polypeptides. In a therapeutic embodiment, fusion polypeptide expressed in a retroviral expression vector is used to infect bone marrow cells that are subsequently transplanted into

irradiated recipients. The pathology of the subject recipient is then examined after sufficient time has passed (e.g., six (6) weeks).

[0048] Expressing the polypeptide in host bacteria with an impaired capacity to proteolytically cleave the recombinant polypeptide is often used to maximize recombinant polypeptide expression (Gottesman, S., *Gene Expression Technology: Methods in Enzymology*, Academic Press, San Diego, California 185: 119-128 (1990)). Another strategy is to alter the nucleotide sequence of the nucleic acid to be inserted into an expression vector so that the individual codons for each amino acid are those preferentially utilized in *E. coli* (Wada *et al.*, *Nucleic Acids Res.* 20: 2111-2118 (1992)). Such alteration of nucleotide sequences can be carried out by standard DNA synthesis techniques.

[0049] When used in mammalian cells, the expression vector's control functions are often provided by viral regulatory elements. For example, commonly used promoters are derived from polyoma, Adenovirus 2, cytomegalovirus and Simian Virus 40. Recombinant mammalian expression vectors are often capable of directing expression of the nucleic acid in a particular cell type (e.g., tissue-specific regulatory elements are used to express the nucleic acid). Non-limiting examples of suitable tissue-specific promoters include an albumin promoter (liver-specific; Pinkert *et al.*, *Genes Dev.* 1: 268-277 (1987)), lymphoid-specific promoters (Calame & Eaton, *Adv. Immunol.* 43: 235-275 (1988)), promoters of T cell receptors (Winoto & Baltimore, *EMBO J.* 8: 729-733 (1989)) promoters of immunoglobulins (Banerji *et al.*, *Cell* 33: 729-740 (1983); Queen & Baltimore, *Cell* 33: 741-748 (1983)), neuron-specific promoters (e.g., the neurofilament promoter; Byrne & Ruddle, *Proc. Natl. Acad. Sci. USA* 86: 5473-5477 (1989)), pancreas-specific promoters (Edlund *et al.*, *Science* 230: 912-916 (1985)), and mammary gland-specific promoters (e.g., milk whey promoter; U.S. Patent No. 4,873,316 and European Application Publication No. 264,166). Developmentally-regulated promoters are sometimes utilized, for example, the murine hox promoters (Kessel & Gruss, *Science* 249: 374-379 (1990)) and the α -fetopolypeptide promoter (Campes & Tilghman, *Genes Dev.* 3: 537-546 (1989)).

[0050] A *TM7SF3* nucleic acid also may be cloned into an expression vector in an antisense orientation. Regulatory sequences (e.g., viral promoters and/or enhancers) operatively linked to a *TM7SF3* nucleic acid cloned in the antisense orientation can be chosen for directing constitutive, tissue specific or cell type specific expression of antisense RNA in a variety of cell types. Antisense expression vectors can be in the form of a recombinant plasmid, phagemid or attenuated virus. For a discussion of the regulation of gene expression using antisense genes see, e.g., Weintraub *et al.*, Antisense RNA as a molecular tool for genetic analysis, *Reviews - Trends in Genetics*, Vol. 1(1) (1986).

[0051] Also provided herein are host cells that include a *TM7SF3* nucleotide sequence within a recombinant expression vector or a fragment of such a nucleotide sequence which facilitate homologous recombination into a specific site of the host cell genome. The terms "host cell" and "recombinant host

cell” are used interchangeably herein. Such terms refer not only to the particular subject cell but rather also to the progeny or potential progeny of such a cell. Because certain modifications may occur in succeeding generations due to either mutation or environmental influences, such progeny may not, in fact, be identical to the parent cell, but are still included within the scope of the term as used herein. A host cell can be any prokaryotic or eukaryotic cell. For example, a target polypeptide can be expressed in bacterial cells such as *E. coli*, insect cells, yeast or mammalian cells (such as Chinese hamster ovary cells (CHO) or COS cells). Other suitable host cells are known to those skilled in the art.

[0052] Vectors can be introduced into host cells via conventional transformation or transfection techniques. As used herein, the terms “transformation” and “transfection” are intended to refer to a variety of art-recognized techniques for introducing foreign nucleic acid (*e.g.*, DNA) into a host cell, including calcium phosphate or calcium chloride co-precipitation, transduction/infection, DEAE-dextran-mediated transfection, lipofection, or electroporation.

[0053] A host cell provided herein can be used to produce (*i.e.*, express) a target polypeptide or a substantially identical polypeptide thereof. Accordingly, further provided are methods for producing a target polypeptide using host cells described herein. In one embodiment, the method includes culturing host cells into which a recombinant expression vector encoding a target polypeptide has been introduced in a suitable medium such that a target polypeptide is produced. In another embodiment, the method further includes isolating a target polypeptide from the medium or the host cell.

[0054] Also provided are cells or purified preparations of cells which include a *TM7SF3* transgene, or which otherwise misexpress target polypeptide. Cell preparations can consist of human or non-human cells, *e.g.*, rodent cells, *e.g.*, mouse or rat cells, rabbit cells, or pig cells. In preferred embodiments, the cell or cells include a *TM7SF3* transgene (*e.g.*, a heterologous form of a *TM7SF3* gene, such as a human gene expressed in non-human cells). The transgene can be misexpressed, *e.g.*, overexpressed or underexpressed. In other preferred embodiments, the cell or cells include a gene which misexpress an endogenous target polypeptide (*e.g.*, expression of a gene is disrupted, also known as a knockout). Such cells can serve as a model for studying disorders which are related to mutated or mis-expressed alleles or for use in drug screening. Also provided are human cells (*e.g.*, a hematopoietic stem cells) transfected with a *TM7SF3* nucleic acid.

[0055] Also provided are cells or a purified preparation thereof (*e.g.*, human cells) in which an endogenous *TM7SF3* nucleic acid is under the control of a regulatory sequence that does not normally control the expression of the endogenous gene. The expression characteristics of an endogenous gene within a cell (*e.g.*, a cell line or microorganism) can be modified by inserting a heterologous DNA regulatory element into the genome of the cell such that the inserted regulatory element is operably linked to the corresponding endogenous gene. For example, an endogenous corresponding gene (*e.g.*, a gene which is “transcriptionally silent,” not normally expressed, or expressed only at very low levels)

may be activated by inserting a regulatory element which is capable of promoting the expression of a normally expressed gene product in that cell. Techniques such as targeted homologous recombinations, can be used to insert the heterologous DNA as described in, *e.g.*, Chappel, US 5,272,071; WO 91/06667, published on May 16, 1991.

Transgenic Animals

[0056] Non-human transgenic animals that express a heterologous target polypeptide (*e.g.*, expressed from a *TM7SF3* nucleic acid or substantially identical sequence thereof) can be generated. Such animals are useful for studying the function and/or activity of a target polypeptide and for identifying and/or evaluating modulators of the activity of *TM7SF3* nucleic acids and encoded polypeptides. As used herein, a "transgenic animal" is a non-human animal such as a mammal (*e.g.*, a non-human primate such as chimpanzee, baboon, or macaque; an ungulate such as an equine, bovine, or caprine; or a rodent such as a rat, a mouse, or an Israeli sand rat), a bird (*e.g.*, a chicken or a turkey), an amphibian (*e.g.*, a frog, salamander, or newt), or an insect (*e.g.*, *Drosophila melanogaster*), in which one or more of the cells of the animal includes a transgene. A transgene is exogenous DNA or a rearrangement (*e.g.*, a deletion of endogenous chromosomal DNA) that is often integrated into or occurs in the genome of cells in a transgenic animal. A transgene can direct expression of an encoded gene product in one or more cell types or tissues of the transgenic animal, and other transgenes can reduce expression (*e.g.*, a knockout). Thus, a transgenic animal can be one in which an endogenous nucleic acid homologous to a *TM7SF3* nucleic acid has been altered by homologous recombination between the endogenous gene and an exogenous DNA molecule introduced into a cell of the animal (*e.g.*, an embryonic cell of the animal) prior to development of the animal.

[0057] Intronic sequences and polyadenylation signals can also be included in the transgene to increase expression efficiency of the transgene. One or more tissue-specific regulatory sequences can be operably linked to a *TM7SF3* nucleotide sequence to direct expression of an encoded polypeptide to particular cells. A transgenic founder animal can be identified based upon the presence of a *TM7SF3* nucleotide sequence in its genome and/or expression of encoded mRNA in tissues or cells of the animals. A transgenic founder animal can then be used to breed additional animals carrying the transgene. Moreover, transgenic animals carrying a *TM7SF3* nucleotide sequence can further be bred to other transgenic animals carrying other transgenes.

[0058] Target polypeptides can be expressed in transgenic animals or plants by introducing, for example, a *TM7SF3* nucleic acid into the genome of an animal that encodes the target polypeptide. In preferred embodiments the nucleic acid is placed under the control of a tissue specific promoter, *e.g.*, a milk or egg specific promoter, and recovered from the milk or eggs produced by the animal. Also included is a population of cells from a transgenic animal.

Target Polypeptides

[0059] Also featured herein are isolated target polypeptides, which are encoded by a *TM7SF3* nucleotide sequence (e.g., SEQ ID NO: 1-2), or a substantially identical nucleotide sequence thereof. Examples of *TM7SF3* polypeptides are set forth in SEQ ID NO: 3. The term “polypeptide” as used herein includes proteins and peptides. An “isolated” or “purified” polypeptide or protein is substantially free of cellular material or other contaminating proteins from the cell or tissue source from which the protein is derived, or substantially free from chemical precursors or other chemicals when chemically synthesized. In one embodiment, the language “substantially free” means preparation of a target polypeptide having less than about 30%, 20%, 10% and more preferably 5% (by dry weight), of non-target polypeptide (also referred to herein as a “contaminating protein”), or of chemical precursors or non-target chemicals. When the target polypeptide or a biologically active portion thereof is recombinantly produced, it is also preferably substantially free of culture medium, specifically, where culture medium represents less than about 20%, sometimes less than about 10%, and often less than about 5% of the volume of the polypeptide preparation. Isolated or purified target polypeptide preparations are sometimes 0.01 milligrams or more or 0.1 milligrams or more, and often 1.0 milligrams or more and 10 milligrams or more in dry weight.

[0060] Further included herein are target polypeptide fragments. The polypeptide fragment may be a domain or part of a domain of a target polypeptide. The polypeptide fragment may have increased, decreased or unexpected biological activity. The polypeptide fragment is often 50 or fewer, 100 or fewer, or 200 or fewer amino acids in length, and is sometimes 300, 400, 500, 600, 700, or 900 or fewer amino acids in length.

[0061] Substantially identical target polypeptides may depart from the amino acid sequences of target polypeptides in different manners. For example, conservative amino acid modifications may be introduced at one or more positions in the amino acid sequences of target polypeptides. A “conservative amino acid substitution” is one in which the amino acid is replaced by another amino acid having a similar structure and/or chemical function. Families of amino acid residues having similar structures and functions are well known. These families include amino acids with basic side chains (e.g., lysine, arginine, histidine), acidic side chains (e.g., aspartic acid, glutamic acid), uncharged polar side chains (e.g., glycine, asparagine, glutamine, serine, threonine, tyrosine, cysteine), nonpolar side chains (e.g., alanine, valine, leucine, isoleucine, proline, phenylalanine, methionine, tryptophan), beta-branched side chains (e.g., threonine, valine, isoleucine) and aromatic side chains (e.g., tyrosine, phenylalanine, tryptophan, histidine). Also, essential and non-essential amino acids may be replaced. A “non-essential” amino acid is one that can be altered without abolishing or substantially altering the biological function of a target polypeptide, whereas altering an “essential” amino acid abolishes or substantially alters the biological function of a target polypeptide. Amino acids that are conserved among target polypeptides

are typically essential amino acids. In certain embodiments, the polypeptide includes one or more non-synonymous polymorphic variants associated with osteoarthritis.

[0062] Also, target polypeptides may exist as chimeric or fusion polypeptides. As used herein, a target "chimeric polypeptide" or target "fusion polypeptide" includes a target polypeptide linked to a non-target polypeptide. A "non-target polypeptide" refers to a polypeptide having an amino acid sequence corresponding to a polypeptide which is not substantially identical to the target polypeptide, which includes, for example, a polypeptide that is different from the target polypeptide and derived from the same or a different organism. The target polypeptide in the fusion polypeptide can correspond to an entire or nearly entire target polypeptide or a fragment thereof. The non-target polypeptide can be fused to the N-terminus or C-terminus of the target polypeptide.

[0063] Fusion polypeptides can include a moiety having high affinity for a ligand. For example, the fusion polypeptide can be a GST-target fusion polypeptide in which the target sequences are fused to the C-terminus of the GST sequences, or a polyhistidine-target fusion polypeptide in which the target polypeptide is fused at the N- or C-terminus to a string of histidine residues. Such fusion polypeptides can facilitate purification of recombinant target polypeptide. Expression vectors are commercially available that already encode a fusion moiety (e.g., a GST polypeptide), and a nucleotide sequence in SEQ ID NO: 1-2, or a substantially identical nucleotide sequence thereof, can be cloned into an expression vector such that the fusion moiety is linked in-frame to the target polypeptide. Further, the fusion polypeptide can be a target polypeptide containing a heterologous signal sequence at its N-terminus. In certain host cells (e.g., mammalian host cells), expression, secretion, cellular internalization, and cellular localization of a target polypeptide can be increased through use of a heterologous signal sequence. Fusion polypeptides can also include all or a part of a serum polypeptide (e.g., an IgG constant region or human serum albumin).

[0064] Target polypeptides can be incorporated into pharmaceutical compositions and administered to a subject *in vivo*. Administration of these target polypeptides can be used to affect the bioavailability of a substrate of the target polypeptide and may effectively increase target polypeptide biological activity in a cell. Target fusion polypeptides may be useful therapeutically for the treatment of disorders caused by, for example, (i) aberrant modification or mutation of a gene encoding a target polypeptide; (ii) mis-regulation of the gene encoding the target polypeptide; and (iii) aberrant post-translational modification of a target polypeptide. Also, target polypeptides can be used as immunogens to produce anti-target antibodies in a subject, to purify target polypeptide ligands or binding partners, and in screening assays to identify molecules which inhibit or enhance the interaction of a target polypeptide with a substrate.

[0065] In addition, polypeptides can be chemically synthesized using techniques known in the art (See, e.g., Creighton, 1983 Proteins. New York, N.Y.: W. H. Freeman and Company; and Hunkapiller et al., (1984) Nature July 12 -18;310(5973):105-11). For example, a relative short fragment can be

synthesized by use of a peptide synthesizer. Furthermore, if desired, non-classical amino acids or chemical amino acid analogs can be introduced as a substitution or addition into the fragment sequence. Non-classical amino acids include, but are not limited to, to the D-isomers of the common amino acids, 2,4-diaminobutyric acid, α -amino isobutyric acid, 4-aminobutyric acid, Abu, 2-amino butyric acid, g-Abu, e-Ahx, 6-amino hexanoic acid, Aib, 2-amino isobutyric acid, 3-amino propionic acid, ornithine, norleucine, norvaline, hydroxyproline, sarcosine, citrulline, homocitrulline, cysteic acid, t-butylglycine, t-butylalanine, phenylglycine, cyclohexylalanine, b-alanine, fluoroamino acids, designer amino acids such as b-methyl amino acids, Ca-methyl amino acids, Na-methyl amino acids, and amino acid analogs in general. Furthermore, the amino acid can be D (dextrorotary) or L (levorotary).

[0066] Polypeptides and polypeptide fragments sometimes are differentially modified during or after translation, *e.g.*, by glycosylation, acetylation, phosphorylation, amidation, derivatization by known protecting/blocking groups, proteolytic cleavage, linkage to an antibody molecule or other cellular ligand, etc. Any of numerous chemical modifications may be carried out by known techniques, including but not limited, to specific chemical cleavage by cyanogen bromide, trypsin, chymotrypsin, papain, V8 protease, NaBH₄; acetylation, formylation, oxidation, reduction; metabolic synthesis in the presence of tunicamycin; and the like. Additional post-translational modifications include, for example, N-linked or O-linked carbohydrate chains, processing of N-terminal or C-terminal ends), attachment of chemical moieties to the amino acid backbone, chemical modifications of N-linked or O-linked carbohydrate chains, and addition or deletion of an N-terminal methionine residue as a result of prokaryotic host cell expression. The polypeptide fragments may also be modified with a detectable label, such as an enzymatic, fluorescent, isotopic or affinity label to allow for detection and isolation of the polypeptide.

[0067] Also provided are chemically modified derivatives of polypeptides that can provide additional advantages such as increased solubility, stability and circulating time of the polypeptide, or decreased immunogenicity (*see e.g.*, U.S. Pat. No: 4,179,337. The chemical moieties for derivatization may be selected from water soluble polymers such as polyethylene glycol, ethylene glycol/propylene glycol copolymers, carboxymethylcellulose, dextran, polyvinyl alcohol and the like. The polypeptides may be modified at random positions within the molecule, or at predetermined positions within the molecule and may include one, two, three or more attached chemical moieties.

[0068] The polymer may be of any molecular weight, and may be branched or unbranched. For polyethylene glycol, the preferred molecular weight is between about 1 kDa and about 100 kDa (the term "about" indicating that in preparations of polyethylene glycol, some molecules will weigh more, some less, than the stated molecular weight) for ease in handling and manufacturing. Other sizes may be used, depending on the desired therapeutic profile (*e.g.*, the duration of sustained release desired, the effects, if any on biological activity, the ease in handling, the degree or lack of antigenicity and other known effects of the polyethylene glycol to a therapeutic protein or analog).

[0069] The polymers should be attached to the polypeptide with consideration of effects on functional or antigenic domains of the polypeptide. There are a number of attachment methods available to those skilled in the art (e.g., EP 0 401 384 (coupling PEG to G-CSF) and Malik et al. (1992) Exp Hematol. September;20(8):1028-35 (pegylation of GM-CSF using tresyl chloride)). For example, polyethylene glycol may be covalently bound through amino acid residues via a reactive group, such as a free amino or carboxyl group. Reactive groups are those to which an activated polyethylene glycol molecule may be bound. The amino acid residues having a free amino group may include lysine residues and the N-terminal amino acid residues; those having a free carboxyl group may include aspartic acid residues, glutamic acid residues and the C-terminal amino acid residue. Sulfhydryl groups may also be used as a reactive group for attaching the polyethylene glycol molecules. For therapeutic purposes, the attachment sometimes is at an amino group, such as attachment at the N-terminus or lysine group.

[0070] Proteins can be chemically modified at the N-terminus. Using polyethylene glycol as an illustration of such a composition, one may select from a variety of polyethylene glycol molecules (by molecular weight, branching, and the like), the proportion of polyethylene glycol molecules to protein (polypeptide) molecules in the reaction mix, the type of pegylation reaction to be performed, and the method of obtaining the selected N-terminally pegylated protein. The method of obtaining the N-terminally pegylated preparation (i.e., separating this moiety from other monopegylated moieties if necessary) may be by purification of the N-terminally pegylated material from a population of pegylated protein molecules. Selective proteins chemically modified at the N-terminus may be accomplished by reductive alkylation, which exploits differential reactivity of different types of primary amino groups (lysine versus the N-terminal) available for derivatization in a particular protein. Under the appropriate reaction conditions, substantially selective derivatization of the protein at the N-terminus with a carbonyl group containing polymer is achieved.

Substantially Identical Nucleic Acids and Polypeptides

[0071] Nucleotide sequences and polypeptide sequences that are substantially identical to a *TM7SF3* nucleotide sequence and the target polypeptide sequences encoded by those nucleotide sequences, respectively, are included herein. The term "substantially identical" as used herein refers to two or more nucleic acids or polypeptides sharing one or more identical nucleotide sequences or polypeptide sequences, respectively. Included are nucleotide sequences or polypeptide sequences that are 55% or more, 60% or more, 65% or more, 70% or more, 75% or more, 80% or more, 85% or more, 90% or more, 95% or more (each often within a 1%, 2%, 3% or 4% variability) identical to a *TM7SF3* nucleotide sequence or the encoded target polypeptide amino acid sequences. One test for determining whether two nucleic acids are substantially identical is to determine the percent of identical nucleotide sequences or polypeptide sequences shared between the nucleic acids or polypeptides.

[0072] Calculations of sequence identity are often performed as follows. Sequences are aligned for optimal comparison purposes (e.g., gaps can be introduced in one or both of a first and a second amino acid or nucleic acid sequence for optimal alignment and non-homologous sequences can be disregarded for comparison purposes). The length of a reference sequence aligned for comparison purposes is sometimes 30% or more, 40% or more, 50% or more, often 60% or more, and more often 70% or more, 80% or more, 90% or more, or 100% of the length of the reference sequence. The nucleotides or amino acids at corresponding nucleotide or polypeptide positions, respectively, are then compared among the two sequences. When a position in the first sequence is occupied by the same nucleotide or amino acid as the corresponding position in the second sequence, the nucleotides or amino acids are deemed to be identical at that position. The percent identity between the two sequences is a function of the number of identical positions shared by the sequences, taking into account the number of gaps, and the length of each gap, introduced for optimal alignment of the two sequences.

[0073] Comparison of sequences and determination of percent identity between two sequences can be accomplished using a mathematical algorithm. Percent identity between two amino acid or nucleotide sequences can be determined using the algorithm of Meyers & Miller, *CABIOS* 4: 11-17 (1989), which has been incorporated into the ALIGN program (version 2.0), using a PAM120 weight residue table, a gap length penalty of 12 and a gap penalty of 4. Also, percent identity between two amino acid sequences can be determined using the Needleman & Wunsch, *J. Mol. Biol.* 48: 444-453 (1970) algorithm which has been incorporated into the GAP program in the GCG software package (available at the http address www.gcg.com), using either a Blossum 62 matrix or a PAM250 matrix, and a gap weight of 16, 14, 12, 10, 8, 6, or 4 and a length weight of 1, 2, 3, 4, 5, or 6. Percent identity between two nucleotide sequences can be determined using the GAP program in the GCG software package (available at http address www.gcg.com), using a NWSgapdna.CMP matrix and a gap weight of 40, 50, 60, 70, or 80 and a length weight of 1, 2, 3, 4, 5, or 6. A set of parameters often used is a Blossum 62 scoring matrix with a gap open penalty of 12, a gap extend penalty of 4, and a frameshift gap penalty of 5.

[0074] Another manner for determining if two nucleic acids are substantially identical is to assess whether a polynucleotide homologous to one nucleic acid will hybridize to the other nucleic acid under stringent conditions. As use herein, the term "stringent conditions" refers to conditions for hybridization and washing. Stringent conditions are known to those skilled in the art and can be found in *Current Protocols in Molecular Biology*, John Wiley & Sons, N.Y., 6.3.1-6.3.6 (1989). Aqueous and non-aqueous methods are described in that reference and either can be used. An example of stringent hybridization conditions is hybridization in 6X sodium chloride/sodium citrate (SSC) at about 45°C, followed by one or more washes in 0.2X SSC, 0.1% SDS at 50°C. Another example of stringent hybridization conditions are hybridization in 6X sodium chloride/sodium citrate (SSC) at about 45°C,

followed by one or more washes in 0.2X SSC, 0.1% SDS at 55°C. A further example of stringent hybridization conditions is hybridization in 6X sodium chloride/sodium citrate (SSC) at about 45°C, followed by one or more washes in 0.2X SSC, 0.1% SDS at 60°C. Often, stringent hybridization conditions are hybridization in 6X sodium chloride/sodium citrate (SSC) at about 45°C, followed by one or more washes in 0.2X SSC, 0.1% SDS at 65°C. More often, stringency conditions are 0.5M sodium phosphate, 7% SDS at 65°C, followed by one or more washes at 0.2X SSC, 1% SDS at 65°C.

[0075] An example of a substantially identical nucleotide sequence to a nucleotide sequence in SEQ ID NO: 1-2 is one that has a different nucleotide sequence but still encodes the same polypeptide sequence encoded by the nucleotide sequence in SEQ ID NO: 1-2. Another example is a nucleotide sequence that encodes a polypeptide having a polypeptide sequence that is more than 70% or more identical to, sometimes more than 75% or more, 80% or more, or 85% or more identical to, and often more than 90% or more and 95% or more identical to a polypeptide sequence encoded by a nucleotide sequence in SEQ ID NO: 1-2.

[0076] Nucleotide sequences in SEQ ID NO: 1-2 and amino acid sequences of encoded polypeptides can be used as "query sequences" to perform a search against public databases to identify other family members or related sequences, for example. Such searches can be performed using the NBLAST and XBLAST programs (version 2.0) of Altschul *et al.*, *J. Mol. Biol.* 215: 403-10 (1990). BLAST nucleotide searches can be performed with the NBLAST program, score = 100, wordlength = 12 to obtain nucleotide sequences homologous to nucleotide sequences in SEQ ID NO: 1-2. BLAST polypeptide searches can be performed with the XBLAST program, score = 50, wordlength = 3 to obtain amino acid sequences homologous to polypeptides encoded by the nucleotide sequences of SEQ ID NO: 1-2. To obtain gapped alignments for comparison purposes, Gapped BLAST can be utilized as described in Altschul *et al.*, *Nucleic Acids Res.* 25(17): 3389-3402 (1997). When utilizing BLAST and Gapped BLAST programs, default parameters of the respective programs (*e.g.*, XBLAST and NBLAST) can be used (*see* the http address www.ncbi.nlm.nih.gov).

[0077] A nucleic acid that is substantially identical to a nucleotide sequence in SEQ ID NO: 1-2 may include polymorphic sites at positions equivalent to those described herein when the sequences are aligned. For example, using the alignment procedures described herein, SNPs in a sequence substantially identical to a sequence in SEQ ID NO: 1-2 can be identified at nucleotide positions that match (*i.e.*, align) with nucleotides at SNP positions in each nucleotide sequence in SEQ ID NO: 1-2. Also, where a polymorphic variation results in an insertion or deletion, insertion or deletion of a nucleotide sequence from a reference sequence can change the relative positions of other polymorphic sites in the nucleotide sequence.

[0078] Substantially identical nucleotide and polypeptide sequences include those that are naturally occurring, such as allelic variants (same locus), splice variants, homologs (different locus), and orthologs (different organism) or can be non-naturally occurring. Non-naturally occurring variants can be generated by mutagenesis techniques, including those applied to polynucleotides, cells, or organisms. The variants can contain nucleotide substitutions, deletions, inversions and insertions. Variation can occur in either or both the coding and non-coding regions. The variations can produce both conservative and non-conservative amino acid substitutions (as compared in the encoded product). Orthologs, homologs, allelic variants, and splice variants can be identified using methods known in the art. These variants normally comprise a nucleotide sequence encoding a polypeptide that is 50% or more, about 55% or more, often about 70-75% or more or about 80-85% or more, and sometimes about 90-95% or more identical to the amino acid sequences of target polypeptides or a fragment thereof. Such nucleic acid molecules can readily be identified as being able to hybridize under stringent conditions to a nucleotide sequence in SEQ ID NO: 1-2 or a fragment of this sequence. Nucleic acid molecules corresponding to orthologs, homologs, and allelic variants of a nucleotide sequence in SEQ ID NO: 1-2 can further be identified by mapping the sequence to the same chromosome or locus as the nucleotide sequence in SEQ ID NO: 1-2.

[0079] Also, substantially identical nucleotide sequences may include codons that are altered with respect to the naturally occurring sequence for enhancing expression of a target polypeptide in a particular expression system. For example, the nucleic acid can be one in which one or more codons are altered, and often 10% or more or 20% or more of the codons are altered for optimized expression in bacteria (*e.g.*, *E. coli*), yeast (*e.g.*, *S. cerevisiae*), human (*e.g.*, 293 cells), insect, or rodent (*e.g.*, hamster) cells.

Methods for Identifying Risk of Osteoarthritis

[0080] Methods for prognosing and diagnosing osteoarthritis are included herein. These methods include detecting the presence or absence of one or more polymorphic variations in a nucleotide sequence associated with osteoarthritis, such as variants in or around the loci set forth herein, or a substantially identical sequence thereof, in a sample from a subject, where the presence of a polymorphic variant described herein is indicative of a risk of osteoarthritis. Determining a risk of osteoarthritis sometimes refers to determining whether an individual is at an increased risk of osteoarthritis (*e.g.*, intermediate risk or higher risk).

[0081] Thus, featured herein is a method for identifying a subject who is at risk of osteoarthritis, which comprises detecting an aberration associated with osteoarthritis in a nucleic acid sample from the subject. An embodiment is a method for detecting a risk of osteoarthritis in a subject, which comprises detecting the presence or absence of a polymorphic variation associated with osteoarthritis at a

polymorphic site in a nucleotide sequence in a nucleic acid sample from a subject, where the nucleotide sequence comprises a polynucleotide sequence selected from the group consisting of: (a) a nucleotide sequence of SEQ ID NO: 1-2; (b) a nucleotide sequence which encodes a polypeptide consisting of an amino acid sequence encoded by a nucleotide sequence of SEQ ID NO: 1-2; (c) a nucleotide sequence which encodes a polypeptide that is 90% or more identical to an amino acid sequence encoded by a nucleotide sequence of SEQ ID NO: 1-2, or a nucleotide sequence about 90% or more identical to a nucleotide sequence of SEQ ID NO: 1-2; and (d) a fragment of a nucleotide sequence of (a), (b), or (c) comprising the polymorphic site; whereby the presence of the polymorphic variation is indicative of a predisposition to osteoarthritis in the subject. In certain embodiments, polymorphic variants at the positions described herein are detected for determining a risk of osteoarthritis, and polymorphic variants at positions in linkage disequilibrium with these positions are detected for determining a risk of osteoarthritis. As used herein, "SEQ ID NO: 1-2" refers to individual sequences in SEQ ID NO: 1 or 2, each sequence being separately applicable to embodiments described herein.

[0082] Risk of osteoarthritis sometimes is expressed as a probability, such as an odds ratio, percentage, or risk factor. Risk often is based upon the presence or absence of one or more polymorphic variants described herein, and also may be based in part upon phenotypic traits of the individual being tested. Methods for calculating risk based upon patient data are well known (*see, e.g., Agresti, Categorical Data Analysis*, 2nd Ed. 2002. Wiley). Allelotyping and genotyping analyses may be carried out in populations other than those exemplified herein to enhance the predictive power of the prognostic method. These further analyses are executed in view of the exemplified procedures described herein, and may be based upon the same polymorphic variations or additional polymorphic variations.

[0083] In certain embodiments, determining the presence of a combination of two or more polymorphic variants associated with osteoarthritis in one or more genetic loci (e.g., one or more genes) of the sample is determined to identify, quantify and/or estimate, risk of osteoarthritis. The risk often is the probability of having or developing osteoarthritis. The risk sometimes is expressed as a relative risk with respect to a population average risk of osteoarthritis, and sometimes is expressed as a relative risk with respect to the lowest risk group. Such relative risk assessments often are based upon penetrance values determined by statistical methods, and are particularly useful to clinicians and insurance companies for assessing risk of osteoarthritis (e.g., a clinician can target appropriate detection, prevention and therapeutic regimens to a patient after determining the patient's risk of osteoarthritis, and an insurance company can fine tune actuarial tables based upon population genotype assessments of osteoarthritis risk). Risk of osteoarthritis sometimes is expressed as an odds ratio, which is the odds of a particular person having a genotype has or will develop osteoarthritis with respect to another genotype group (e.g., the most disease protective genotype or population average). In related embodiments, the determination is utilized to identify a subject at risk of osteoarthritis. In an embodiment, two or more

polymorphic variations are detected in two or more regions in human genomic DNA associated with increased risk of osteoarthritis, such as a locus containing a *TM7SF3*, for example. In certain embodiments, 3 or more, or 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19 or 20 or more polymorphic variants are detected in the sample. In specific embodiments, polymorphic variants are detected in a *TM7SF3* region, for example. In certain embodiments, polymorphic variants are detected at other genetic loci (e.g., the polymorphic variants can be detected in *TM7SF3* in addition to other loci or only in other loci), where the other loci include but are not limited to those described in concurrently-filed patent applications having attorney docket number 524593008700, 524593008800, 524593008900, 524593009000 or 524593009200, which is incorporated herein by reference in its entirety.

[0084] Results from prognostic tests may be combined with other test results to diagnose osteoarthritis. For example, prognostic results may be gathered, a patient sample may be ordered based on a determined predisposition to osteoarthritis, the patient sample is analyzed, and the results of the analysis may be utilized to diagnose osteoarthritis. Also osteoarthritis diagnostic method can be developed from studies used to generate prognostic methods in which populations are stratified into subpopulations having different progressions of osteoarthritis. In another embodiment, prognostic results may be gathered, a patient's risk factors for developing osteoarthritis (e.g., age, weight, race, diet) analyzed, and a patient sample may be ordered based on a determined predisposition to osteoarthritis.

[0085] The nucleic acid sample typically is isolated from a biological sample obtained from a subject. For example, nucleic acid can be isolated from blood, saliva, sputum, urine, cell scrapings, and biopsy tissue. The nucleic acid sample can be isolated from a biological sample using standard techniques, such as the technique described in Example 2. As used herein, the term "subject" refers primarily to humans but also refers to other mammals such as dogs, cats, and ungulates (e.g., cattle, sheep, and swine). Subjects also include avians (e.g., chickens and turkeys), reptiles, and fish (e.g., salmon), as embodiments described herein can be adapted to nucleic acid samples isolated from any of these organisms. The nucleic acid sample may be isolated from the subject and then directly utilized in a method for determining the presence of a polymorphic variant, or alternatively, the sample may be isolated and then stored (e.g., frozen) for a period of time before being subjected to analysis.

[0086] The presence or absence of a polymorphic variant is determined using one or both chromosomal complements represented in the nucleic acid sample. Determining the presence or absence of a polymorphic variant in both chromosomal complements represented in a nucleic acid sample from a subject having a copy of each chromosome is useful for determining the zygosity of an individual for the polymorphic variant (i.e., whether the individual is homozygous or heterozygous for the polymorphic variant). Any oligonucleotide-based diagnostic may be utilized to determine whether a sample includes the presence or absence of a polymorphic variant in a sample. For example, primer extension methods, ligase sequence determination methods (e.g., U.S. Pat. Nos. 5,679,524 and 5,952,174, and WO

01/27326), mismatch sequence determination methods (e.g., U.S. Pat. Nos. 5,851,770; 5,958,692; 6,110,684; and 6,183,958), microarray sequence determination methods, restriction fragment length polymorphism (RFLP), single strand conformation polymorphism detection (SSCP) (e.g., U.S. Pat. Nos. 5,891,625 and 6,013,499), PCR-based assays (e.g., TAQMAN[®] PCR System (Applied Biosystems)), and nucleotide sequencing methods may be used.

[0087] Oligonucleotide extension methods typically involve providing a pair of oligonucleotide primers in a polymerase chain reaction (PCR) or in other nucleic acid amplification methods for the purpose of amplifying a region from the nucleic acid sample that comprises the polymorphic variation. One oligonucleotide primer is complementary to a region 3' of the polymorphism and the other is complementary to a region 5' of the polymorphism. A PCR primer pair may be used in methods disclosed in U.S. Pat. Nos. 4,683,195; 4,683,202; 4,965,188; 5,656,493; 5,998,143; 6,140,054; WO 01/27327; and WO 01/27329 for example. PCR primer pairs may also be used in any commercially available machines that perform PCR, such as any of the GENEAMP[®] Systems available from Applied Biosystems. Also, those of ordinary skill in the art will be able to design oligonucleotide primers based upon a *TM7SF3* nucleotide sequence using knowledge available in the art.

[0088] Also provided is an extension oligonucleotide that hybridizes to the amplified fragment adjacent to the polymorphic variation. As used herein, the term "adjacent" refers to the 3' end of the extension oligonucleotide being often 1 nucleotide from the 5' end of the polymorphic site, and sometimes 2, 3, 4, 5, 6, 7, 8, 9, or 10 nucleotides from the 5' end of the polymorphic site, in the nucleic acid when the extension oligonucleotide is hybridized to the nucleic acid. The extension oligonucleotide then is extended by one or more nucleotides, and the number and/or type of nucleotides that are added to the extension oligonucleotide determine whether the polymorphic variant is present. Oligonucleotide extension methods are disclosed, for example, in U.S. Pat. Nos. 4,656,127; 4,851,331; 5,679,524; 5,834,189; 5,876,934; 5,908,755; 5,912,118; 5,976,802; 5,981,186; 6,004,744; 6,013,431; 6,017,702; 6,046,005; 6,087,095; 6,210,891; and WO 01/20039. Oligonucleotide extension methods using mass spectrometry are described, for example, in U.S. Pat. Nos. 5,547,835; 5,605,798; 5,691,141; 5,849,542; 5,869,242; 5,928,906; 6,043,031; and 6,194,144, and a method often utilized is described herein in Example 2.

[0089] A microarray can be utilized for determining whether a polymorphic variant is present or absent in a nucleic acid sample. A microarray may include any oligonucleotides described herein, and methods for making and using oligonucleotide microarrays suitable for diagnostic use are disclosed in U.S. Pat. Nos. 5,492,806; 5,525,464; 5,589,330; 5,695,940; 5,849,483; 6,018,041; 6,045,996; 6,136,541; 6,142,681; 6,156,501; 6,197,506; 6,223,127; 6,225,625; 6,229,911; 6,239,273; WO 00/52625; WO 01/25485; and WO 01/29259. The microarray typically comprises a solid support and the oligonucleotides may be linked to this solid support by covalent bonds or by non-covalent interactions.

The oligonucleotides may also be linked to the solid support directly or by a spacer molecule. A microarray may comprise one or more oligonucleotides complementary to a polymorphic site set forth herein.

[0090] A kit also may be utilized for determining whether a polymorphic variant is present or absent in a nucleic acid sample. A kit often comprises one or more pairs of oligonucleotide primers useful for amplifying a fragment of a nucleotide sequence of SEQ ID NO: 1-2 or a substantially identical sequence thereof, where the fragment includes a polymorphic site. The kit sometimes comprises a polymerizing agent, for example, a thermostable nucleic acid polymerase such as one disclosed in U.S. Pat. Nos. 4,889,818 or 6,077,664. Also, the kit often comprises an elongation oligonucleotide that hybridizes to a *TM7SF3* nucleotide sequence in a nucleic acid sample adjacent to the polymorphic site. Where the kit includes an elongation oligonucleotide, it also often comprises chain elongating nucleotides, such as dATP, dTTP, dGTP, dCTP, and dITP, including analogs of dATP, dTTP, dGTP, dCTP and dITP, provided that such analogs are substrates for a thermostable nucleic acid polymerase and can be incorporated into a nucleic acid chain elongated from the extension oligonucleotide. Along with chain elongating nucleotides would be one or more chain terminating nucleotides such as ddATP, ddTTP, ddGTP, ddCTP, and the like. In an embodiment, the kit comprises one or more oligonucleotide primer pairs, a polymerizing agent, chain elongating nucleotides, at least one elongation oligonucleotide, and one or more chain terminating nucleotides. Kits optionally include buffers, vials, microtiter plates, and instructions for use.

[0091] An individual identified as being at risk of osteoarthritis may be heterozygous or homozygous with respect to the allele associated with a higher risk of osteoarthritis. A subject homozygous for an allele associated with an increased risk of osteoarthritis is at a comparatively high risk of osteoarthritis, a subject heterozygous for an allele associated with an increased risk of osteoarthritis is at a comparatively intermediate risk of osteoarthritis, and a subject homozygous for an allele associated with a decreased risk of osteoarthritis is at a comparatively low risk of osteoarthritis. A genotype may be assessed for a complementary strand, such that the complementary nucleotide at a particular position is detected.

[0092] Also featured are methods for determining risk of osteoarthritis and/or identifying a subject at risk of osteoarthritis by contacting a polypeptide or protein encoded by a *TM7SF3* nucleotide sequence from a subject with an antibody that specifically binds to an epitope associated with increased risk of osteoarthritis in the polypeptide.

Applications of Prognostic and Diagnostic Results to Pharmacogenomic Methods

[0093] Pharmacogenomics is a discipline that involves tailoring a treatment for a subject according to the subject's genotype as a particular treatment regimen may exert a differential effect depending upon

the subject's genotype. For example, based upon the outcome of a prognostic test described herein, a clinician or physician may target pertinent information and preventative or therapeutic treatments to a subject who would be benefited by the information or treatment and avoid directing such information and treatments to a subject who would not be benefited (e.g., the treatment has no therapeutic effect and/or the subject experiences adverse side effects).

[0094] The following is an example of a pharmacogenomic embodiment. A particular treatment regimen can exert a differential effect depending upon the subject's genotype. Where a candidate therapeutic exhibits a significant interaction with a major allele and a comparatively weak interaction with a minor allele (e.g., an order of magnitude or greater difference in the interaction), such a therapeutic typically would not be administered to a subject genotyped as being homozygous for the minor allele, and sometimes not administered to a subject genotyped as being heterozygous for the minor allele. In another example, where a candidate therapeutic is not significantly toxic when administered to subjects who are homozygous for a major allele but is comparatively toxic when administered to subjects heterozygous or homozygous for a minor allele, the candidate therapeutic is not typically administered to subjects who are genotyped as being heterozygous or homozygous with respect to the minor allele.

[0095] The methods described herein are applicable to pharmacogenomic methods for preventing, alleviating or treating osteoarthritis. For example, a nucleic acid sample from an individual may be subjected to a prognostic test described herein. Where one or more polymorphic variations associated with increased risk of osteoarthritis are identified in a subject, information for preventing or treating osteoarthritis and/or one or more osteoarthritis treatment regimens then may be prescribed to that subject.

[0096] In certain embodiments, a treatment or preventative regimen is specifically prescribed and/or administered to individuals who will most benefit from it based upon their risk of developing osteoarthritis assessed by the methods described herein. Thus, provided are methods for identifying a subject predisposed to osteoarthritis and then prescribing a therapeutic or preventative regimen to individuals identified as having a predisposition. Thus, certain embodiments are directed to a method for reducing osteoarthritis in a subject, which comprises: detecting the presence or absence of a polymorphic variant associated with osteoarthritis in a nucleotide sequence in a nucleic acid sample from a subject, where the nucleotide sequence comprises a polynucleotide sequence selected from the group consisting of: (a) a nucleotide sequence of SEQ ID NO: 1-2; (b) a nucleotide sequence which encodes a polypeptide consisting of an amino acid sequence encoded by a nucleotide sequence of SEQ ID NO: 1-2; (c) a nucleotide sequence which encodes a polypeptide that is 90% or more identical to an amino acid sequence encoded by a nucleotide sequence of SEQ ID NO: 1-2, or a nucleotide sequence about 90% or more identical to a nucleotide sequence of SEQ ID NO: 1-2; and (d) a fragment of a polynucleotide sequence of (a), (b), or (c); and prescribing or administering a treatment regimen to a subject from whom the sample originated where the presence of a polymorphic variation associated with osteoarthritis is

detected in the nucleotide sequence. In these methods, predisposition results may be utilized in combination with other test results to diagnose osteoarthritis.

[0097] Certain preventative treatments often are prescribed to subjects having a predisposition to osteoarthritis and where the subject is diagnosed with osteoarthritis or is diagnosed as having symptoms indicative of an early stage of osteoarthritis. The treatment sometimes is preventative (e.g., is prescribed or administered to reduce the probability that osteoarthritis arises or progresses), sometimes is therapeutic, and sometimes delays, alleviates or halts the progression of osteoarthritis. Any known preventative or therapeutic treatment for alleviating or preventing the occurrence of osteoarthritis is prescribed and/or administered. For example, the treatment often is directed to decreasing pain and improving joint movement. Examples of OA treatments include exercises to keep joints flexible and improve muscle strength. Different medications to control pain, including corticosteroids and nonsteroidal anti-inflammatory drugs (NSAIDs, e.g., Voltaren); cyclooxygenase-2 (COX-2) inhibitors (e.g., Celebrex, Vioxx, Mobic, and Bextra); monoclonal antibodies (e.g., Remicade); tumor necrosis factor inhibitors (e.g., Enbrel); or injections of glucocorticoids, hyaluronic acid or chondroitin sulfate into joints that are inflamed and not responsive to NSAIDs. Orally administered chondroitin sulfate also may be used as a therapeutic, as it may increase hyaluronic acid levels and viscosity of synovial fluid, and decrease collagenase levels in synovial fluid. Also, glucosamine can serve as an OA therapeutic as delivering it into joints may inhibit enzymes involved in cartilage degradation and enhance the production of hyaluronic acid. For mild pain without inflammation, acetaminophen may be used. Other treatments include: heat/cold therapy for temporary pain relief; joint protection to prevent strain or stress on painful joints; surgery to relieve chronic pain in damaged joints; and weight control to prevent extra stress on weight-bearing joints.

[0098] As therapeutic approaches for treating osteoarthritis continue to evolve and improve, the goal of treatments for osteoarthritis related disorders is to intervene even before clinical signs first manifest. Thus, genetic markers associated with susceptibility to osteoarthritis prove useful for early diagnosis, prevention and treatment of osteoarthritis.

[0099] As osteoarthritis preventative and treatment information can be specifically targeted to subjects in need thereof (e.g., those at risk of developing osteoarthritis or those in an early stage of osteoarthritis), provided herein is a method for preventing or reducing the risk of developing osteoarthritis in a subject, which comprises: (a) detecting the presence or absence of a polymorphic variation associated with osteoarthritis at a polymorphic site in a nucleotide sequence in a nucleic acid sample from a subject; (b) identifying a subject with a predisposition to osteoarthritis, whereby the presence of the polymorphic variation is indicative of a predisposition to osteoarthritis in the subject; and (c) if such a predisposition is identified, providing the subject with information about methods or products to prevent or reduce osteoarthritis or to delay the onset of osteoarthritis. Also provided is a

method of targeting information or advertising to a subpopulation of a human population based on the subpopulation being genetically predisposed to a disease or condition, which comprises: (a) detecting the presence or absence of a polymorphic variation associated with osteoarthritis at a polymorphic site in a nucleotide sequence in a nucleic acid sample from a subject; (b) identifying the subpopulation of subjects in which the polymorphic variation is associated with osteoarthritis; and (c) providing information only to the subpopulation of subjects about a particular product which may be obtained and consumed or applied by the subject to help prevent or delay onset of the disease or condition.

[0100] Pharmacogenomics methods also may be used to analyze and predict a response to osteoarthritis treatment or a drug. For example, if pharmacogenomics analysis indicates a likelihood that an individual will respond positively to osteoarthritis treatment with a particular drug, the drug may be administered to the individual. Conversely, if the analysis indicates that an individual is likely to respond negatively to treatment with a particular drug, an alternative course of treatment may be prescribed. A negative response may be defined as either the absence of an efficacious response or the presence of toxic side effects. The response to a therapeutic treatment can be predicted in a background study in which subjects in any of the following populations are genotyped: a population that responds favorably to a treatment regimen, a population that does not respond significantly to a treatment regimen, and a population that responds adversely to a treatment regimen (*e.g.*, exhibits one or more side effects). These populations are provided as examples and other populations and subpopulations may be analyzed. Based upon the results of these analyses, a subject is genotyped to predict whether he or she will respond favorably to a treatment regimen, not respond significantly to a treatment regimen, or respond adversely to a treatment regimen.

[0101] The tests described herein also are applicable to clinical drug trials. One or more polymorphic variants indicative of response to an agent for treating osteoarthritis or to side effects to an agent for treating osteoarthritis may be identified using the methods described herein. Thereafter, potential participants in clinical trials of such an agent may be screened to identify those individuals most likely to respond favorably to the drug and exclude those likely to experience side effects. In that way, the effectiveness of drug treatment may be measured in individuals who respond positively to the drug, without lowering the measurement as a result of the inclusion of individuals who are unlikely to respond positively in the study and without risking undesirable safety problems.

[0102] Thus, another embodiment is a method of selecting an individual for inclusion in a clinical trial of a treatment or drug comprising the steps of: (a) obtaining a nucleic acid sample from an individual; (b) determining the identity of a polymorphic variation which is associated with a positive response to the treatment or the drug, or at least one polymorphic variation which is associated with a negative response to the treatment or the drug in the nucleic acid sample, and (c) including the individual in the clinical trial if the nucleic acid sample contains said polymorphic variation associated with a

positive response to the treatment or the drug or if the nucleic acid sample lacks said polymorphic variation associated with a negative response to the treatment or the drug. In addition, the methods described herein for selecting an individual for inclusion in a clinical trial of a treatment or drug encompass methods with any further limitation described in this disclosure, or those following, specified alone or in any combination. The polymorphic variation may be in a sequence selected individually or in any combination from the group consisting of (i) a nucleotide sequence of SEQ ID NO: 1-2; (ii) a nucleotide sequence which encodes a polypeptide consisting of an amino acid sequence encoded by a nucleotide sequence of SEQ ID NO: 1-2; (iii) a nucleotide sequence which encodes a polypeptide that is 90% or more identical to an amino acid sequence encoded by a nucleotide sequence of SEQ ID NO: 1-2, or a nucleotide sequence about 90% or more identical to a nucleotide sequence of SEQ ID NO: 1-2; and (iv) a fragment of a polynucleotide sequence of (i), (ii), or (iii) comprising the polymorphic site. The including step (c) optionally comprises administering the drug or the treatment to the individual if the nucleic acid sample contains the polymorphic variation associated with a positive response to the treatment or the drug and the nucleic acid sample lacks said biallelic marker associated with a negative response to the treatment or the drug.

[0103] Also provided herein is a method of partnering between a diagnostic/prognostic testing provider and a provider of a consumable product, which comprises: (a) the diagnostic/prognostic testing provider detects the presence or absence of a polymorphic variation associated with osteoarthritis at a polymorphic site in a nucleotide sequence in a nucleic acid sample from a subject; (b) the diagnostic/prognostic testing provider identifies the subpopulation of subjects in which the polymorphic variation is associated with osteoarthritis; (c) the diagnostic/prognostic testing provider forwards information to the subpopulation of subjects about a particular product which may be obtained and consumed or applied by the subject to help prevent or delay onset of the disease or condition; and (d) the provider of a consumable product forwards to the diagnostic test provider a fee every time the diagnostic/prognostic test provider forwards information to the subject as set forth in step (c) above.

Compositions Comprising Osteoarthritis-Directed Molecules

[0104] Featured herein is a composition comprising a cell from a subject having osteoarthritis or at risk of osteoarthritis and one or more molecules specifically directed and targeted to a nucleic acid comprising a *TM7SF3* nucleotide sequence or amino acid sequence. Such directed molecules include, but are not limited to, a compound that binds to a *TM7SF3* nucleotide sequence or amino acid sequence referenced herein; a RNAi or siRNA molecule having a strand complementary or substantially complementary to a *TM7SF3* nucleotide sequence (e.g., hybridizes to a *TM7SF3* nucleotide sequence under conditions of high stringency); an antisense nucleic acid complementary or substantially complementary to an RNA encoded by a *TM7SF3* nucleotide sequence (e.g., hybridizes to a *TM7SF3*

nucleotide sequence under conditions of high stringency); a ribozyme that hybridizes to a *TM7SF3* nucleotide sequence (e.g., hybridizes to a *TM7SF3* nucleotide sequence under conditions of high stringency); a nucleic acid aptamer that specifically binds a polypeptide encoded by *TM7SF3* nucleotide sequence; and an antibody that specifically binds to a polypeptide encoded by *TM7SF3* nucleotide sequence or binds to a nucleic acid having such a nucleotide sequence. In specific embodiments, the osteoarthritis directed molecule interacts with a nucleic acid or polypeptide variant associated with osteoarthritis, such as variants referenced herein. In other embodiments, the osteoarthritis directed molecule interacts with a polypeptide involved in a signal pathway of a polypeptide encoded by a *TM7SF3* nucleotide sequence, or a nucleic acid comprising such a nucleotide sequence.

[0105] Compositions sometimes include an adjuvant known to stimulate an immune response, and in certain embodiments, an adjuvant that stimulates a T-cell lymphocyte response. Adjuvants are known, including but not limited to an aluminum adjuvant (e.g., aluminum hydroxide); a cytokine adjuvant or adjuvant that stimulates a cytokine response (e.g., interleukin (IL)-12 and/or gamma-interferon cytokines); a Freund-type mineral oil adjuvant emulsion (e.g., Freund's complete or incomplete adjuvant); a synthetic lipid compound; a copolymer adjuvant (e.g., TitreMax); a saponin; Quil A; a liposome; an oil-in-water emulsion (e.g., an emulsion stabilized by Tween 80 and pluronic polyoxyethylene/polyoxypropylene block copolymer (Syntex Adjuvant Formulation); TitreMax; detoxified endotoxin (MPL) and mycobacterial cell wall components (TDW, CWS) in 2% squalene (Ribi Adjuvant System)); a muramyl dipeptide; an immune-stimulating complex (ISCOM, e.g., an Ag-modified saponin/cholesterol micelle that forms stable cage-like structure); an aqueous phase adjuvant that does not have a depot effect (e.g., Gerbu adjuvant); a carbohydrate polymer (e.g., AdjuPrime); L-tyrosine; a manide-oleate compound (e.g., Montanide); an ethylene-vinyl acetate copolymer (e.g., Elvax 40W1,2); or lipid A, for example. Such compositions are useful for generating an immune response against osteoarthritis directed molecule (e.g., an HLA-binding subsequence within a polypeptide encoded by a *TM7SF3* nucleotide sequence). In such methods, a peptide having an amino acid subsequence of a polypeptide encoded by a *TM7SF3* nucleotide sequence is delivered to a subject, where the subsequence binds to an HLA molecule and induces a CTL lymphocyte response. The peptide sometimes is delivered to the subject as an isolated peptide or as a minigene in a plasmid that encodes the peptide. Methods for identifying HLA-binding subsequences in such polypeptides are known (see e.g., publication WO02/20616 and PCT application US98/01373 for methods of identifying such sequences).

[0106] The cell may be in a group of cells cultured *in vitro* or in a tissue maintained *in vitro* or present in an animal *in vivo* (e.g., a rat, mouse, ape or human). In certain embodiments, a composition comprises a component from a cell such as a nucleic acid molecule (e.g., genomic DNA), a protein mixture or isolated protein, for example. The aforementioned compositions have utility in diagnostic,

prognostic and pharmacogenomic methods described previously and in therapeutics described hereafter. Certain osteoarthritis directed molecules are described in greater detail below.

Compounds

[0107] Compounds can be obtained using any of the numerous approaches in combinatorial library methods known in the art, including: biological libraries; peptoid libraries (libraries of molecules having the functionalities of peptides, but with a novel, non-peptide backbone which are resistant to enzymatic degradation but which nevertheless remain bioactive (see, e.g., Zuckermann et al., *J. Med. Chem.* 37: 2678-85 (1994)); spatially addressable parallel solid phase or solution phase libraries; synthetic library methods requiring deconvolution; "one-bead one-compound" library methods; and synthetic library methods using affinity chromatography selection. Biological library and peptoid library approaches are typically limited to peptide libraries, while the other approaches are applicable to peptide, non-peptide oligomer or small molecule libraries of compounds (Lam, *Anticancer Drug Des.* 12: 145, (1997)). Examples of methods for synthesizing molecular libraries are described, for example, in DeWitt et al., *Proc. Natl. Acad. Sci. U.S.A.* 90: 6909 (1993); Erb et al., *Proc. Natl. Acad. Sci. USA* 91: 11422 (1994); Zuckermann et al., *J. Med. Chem.* 37: 2678 (1994); Cho et al., *Science* 261: 1303 (1993); Carrell et al., *Angew. Chem. Int. Ed. Engl.* 33: 2059 (1994); Carrell et al., *Angew. Chem. Int. Ed. Engl.* 33: 2061 (1994); and in Gallop et al., *J. Med. Chem.* 37: 1233 (1994).

[0108] Libraries of compounds may be presented in solution (e.g., Houghten, *Biotechniques* 13: 412-421 (1992)), or on beads (Lam, *Nature* 354: 82-84 (1991)), chips (Fodor, *Nature* 364: 555-556 (1993)), bacteria or spores (Ladner, United States Patent No. 5,223,409), plasmids (Cull et al., *Proc. Natl. Acad. Sci. USA* 89: 1865-1869 (1992)) or on phage (Scott and Smith, *Science* 249: 386-390 (1990); Devlin, *Science* 249: 404-406 (1990); Cwirla et al., *Proc. Natl. Acad. Sci.* 87: 6378-6382 (1990); Felici, *J. Mol. Biol.* 222: 301-310 (1991); Ladner *supra.*).

[0109] A compound sometimes alters expression and sometimes alters activity of a polypeptide target and may be a small molecule. Small molecules include, but are not limited to, peptides, peptidomimetics (e.g., peptoids), amino acids, amino acid analogs, polynucleotides, polynucleotide analogs, nucleotides, nucleotide analogs, organic or inorganic compounds (i.e., including heteroorganic and organometallic compounds) having a molecular weight less than about 10,000 grams per mole, organic or inorganic compounds having a molecular weight less than about 5,000 grams per mole, organic or inorganic compounds having a molecular weight less than about 1,000 grams per mole, organic or inorganic compounds having a molecular weight less than about 500 grams per mole, and salts, esters, and other pharmaceutically acceptable forms of such compounds.

Antisense Nucleic Acid Molecules, Ribozymes, RNAi, siRNA and Modified Nucleic Acid Molecules

[0110] An “antisense” nucleic acid refers to a nucleotide sequence complementary to a “sense” nucleic acid encoding a polypeptide, e.g., complementary to the coding strand of a double-stranded cDNA molecule or complementary to an mRNA sequence. The antisense nucleic acid can be complementary to an entire coding strand, or to a portion thereof or a substantially identical sequence thereof. In another embodiment, the antisense nucleic acid molecule is antisense to a “noncoding region” of the coding strand of a nucleotide sequence (e.g., 5' and 3' untranslated regions in SEQ ID NO: 1).

[0111] An antisense nucleic acid can be designed such that it is complementary to the entire coding region of an mRNA encoded by a nucleotide sequence (e.g., SEQ ID NO: 1), and often the antisense nucleic acid is an oligonucleotide antisense to only a portion of a coding or noncoding region of the mRNA. For example, the antisense oligonucleotide can be complementary to the region surrounding the translation start site of the mRNA, e.g., between the -10 and +10 regions of the target gene nucleotide sequence of interest. An antisense oligonucleotide can be, for example, about 7, 10, 15, 20, 25, 30, 35, 40, 45, 50, 55, 60, 65, 70, 75, 80, or more nucleotides in length. The antisense nucleic acids, which include the ribozymes described hereafter, can be designed to target a *TM7SF3* nucleotide sequence, often a variant associated with osteoarthritis, or a substantially identical sequence thereof. Among the variants, minor alleles and major alleles can be targeted, and those associated with a higher risk of osteoarthritis are often designed, tested, and administered to subjects.

[0112] An antisense nucleic acid can be constructed using chemical synthesis and enzymatic ligation reactions using standard procedures. For example, an antisense nucleic acid (e.g., an antisense oligonucleotide) can be chemically synthesized using naturally occurring nucleotides or variously modified nucleotides designed to increase the biological stability of the molecules or to increase the physical stability of the duplex formed between the antisense and sense nucleic acids, e.g., phosphorothioate derivatives and acridine substituted nucleotides can be used. Antisense nucleic acid also can be produced biologically using an expression vector into which a nucleic acid has been subcloned in an antisense orientation (i.e., RNA transcribed from the inserted nucleic acid will be of an antisense orientation to a target nucleic acid of interest, described further in the following subsection).

[0113] When utilized as therapeutics, antisense nucleic acids typically are administered to a subject (e.g., by direct injection at a tissue site) or generated in situ such that they hybridize with or bind to cellular mRNA and/or genomic DNA encoding a polypeptide and thereby inhibit expression of the polypeptide, for example, by inhibiting transcription and/or translation. Alternatively, antisense nucleic acid molecules can be modified to target selected cells and then are administered systemically. For systemic administration, antisense molecules can be modified such that they specifically bind to receptors or antigens expressed on a selected cell surface, for example, by linking antisense nucleic acid molecules

to peptides or antibodies which bind to cell surface receptors or antigens. Antisense nucleic acid molecules can also be delivered to cells using the vectors described herein. Sufficient intracellular concentrations of antisense molecules are achieved by incorporating a strong promoter, such as a pol II or pol III promoter, in the vector construct.

[0114] Antisense nucleic acid molecules sometimes are alpha-anomeric nucleic acid molecules. An alpha-anomeric nucleic acid molecule forms specific double-stranded hybrids with complementary RNA in which, contrary to the usual beta-units, the strands run parallel to each other (Gaultier et al., *Nucleic Acids. Res.* 15: 6625-6641 (1987)). Antisense nucleic acid molecules can also comprise a 2'-o-methylribonucleotide (Inoue et al., *Nucleic Acids Res.* 15: 6131-6148 (1987)) or a chimeric RNA-DNA analogue (Inoue et al., *FEBS Lett.* 215: 327-330 (1987)). Antisense nucleic acids sometimes are composed of DNA or PNA or any other nucleic acid derivatives described previously.

[0115] In another embodiment, an antisense nucleic acid is a ribozyme. A ribozyme having specificity for a *TM7SF3* nucleotide sequence can include one or more sequences complementary to such a nucleotide sequence, and a sequence having a known catalytic region responsible for mRNA cleavage (see e.g., U.S. Pat. No. 5,093,246 or Haselhoff and Gerlach, *Nature* 334: 585-591 (1988)). For example, a derivative of a *Tetrahymena* L-19 IVS RNA is sometimes utilized in which the nucleotide sequence of the active site is complementary to the nucleotide sequence to be cleaved in a mRNA (see e.g., Cech et al. U.S. Patent No. 4,987,071; and Cech et al. U.S. Patent No. 5,116,742). Also, target mRNA sequences can be used to select a catalytic RNA having a specific ribonuclease activity from a pool of RNA molecules (see e.g., Bartel & Szostak, *Science* 261: 1411-1418 (1993)).

[0116] Osteoarthritis directed molecules include in certain embodiments nucleic acids that can form triple helix structures with a *TM7SF3* nucleotide sequence, or a substantially identical sequence thereof, especially one that includes a regulatory region that controls expression of a polypeptide. Gene expression can be inhibited by targeting nucleotide sequences complementary to the regulatory region of a nucleotide sequence referenced herein or a substantially identical sequence (e.g., promoter and/or enhancers) to form triple helical structures that prevent transcription of a gene in target cells (see e.g., Helene, *Anticancer Drug Des.* 6(6): 569-84 (1991); Helene et al., *Ann. N.Y. Acad. Sci.* 660: 27-36 (1992); and Maher, *Bioassays* 14(12): 807-15 (1992). Potential sequences that can be targeted for triple helix formation can be increased by creating a so-called "switchback" nucleic acid molecule. Switchback molecules are synthesized in an alternating 5'-3', 3'-5' manner, such that they base pair with first one strand of a duplex and then the other, eliminating the necessity for a sizeable stretch of either purines or pyrimidines to be present on one strand of a duplex.

[0117] Osteoarthritis directed molecules include RNAi and siRNA nucleic acids. Gene expression may be inhibited by the introduction of double-stranded RNA (dsRNA), which induces potent and specific gene silencing, a phenomenon called RNA interference or RNAi. See, e.g., Fire et al., US Patent

Number 6,506,559; Tuschl et al. PCT International Publication No. WO 01/75164; Kay et al. PCT International Publication No. WO 03/010180A1; or Bosher JM, Labouesse, Nat Cell Biol 2000 Feb;2(2):E31-6. This process has been improved by decreasing the size of the double-stranded RNA to 20-24 base pairs (to create small-interfering RNAs or siRNAs) that "switched off" genes in mammalian cells without initiating an acute phase response, i.e., a host defense mechanism that often results in cell death (see, e.g., Caplen et al. Proc Natl Acad Sci U S A. 2001 Aug 14;98(17):9742-7 and Elbashir et al. Methods 2002 Feb;26(2):199-213). There is increasing evidence of post-transcriptional gene silencing by RNA interference (RNAi) for inhibiting targeted expression in mammalian cells at the mRNA level, in human cells. There is additional evidence of effective methods for inhibiting the proliferation and migration of tumor cells in human patients, and for inhibiting metastatic cancer development (see, e.g., U.S. Patent Application No. US2001000993183; Caplen et al. Proc Natl Acad Sci U S A; and Abderrahmani et al. Mol Cell Biol 2001 Nov21(21):7256-67).

[0118] An "siRNA" or "RNAi" refers to a nucleic acid that forms a double stranded RNA and has the ability to reduce or inhibit expression of a gene or target gene when the siRNA is delivered to or expressed in the same cell as the gene or target gene. "siRNA" refers to short double-stranded RNA formed by the complementary strands. Complementary portions of the siRNA that hybridize to form the double stranded molecule often have substantial or complete identity to the target molecule sequence. In one embodiment, an siRNA refers to a nucleic acid that has substantial or complete identity to a target gene and forms a double stranded siRNA.

[0119] When designing the siRNA molecules, the targeted region often is selected from a given DNA sequence beginning 50 to 100 nucleotides downstream of the start codon. See, e.g., Elbashir et al., Methods 26:199-213 (2002). Initially, 5' or 3' UTRs and regions nearby the start codon were avoided assuming that UTR-binding proteins and/or translation initiation complexes may interfere with binding of the siRNP or RISC endonuclease complex. Sometimes regions of the target 23 nucleotides in length conforming to the sequence motif AA(N19)TT (N, an nucleotide), and regions with approximately 30% to 70% G/C-content (often about 50% G/C-content) often are selected. If no suitable sequences are found, the search often is extended using the motif NA(N21). The sequence of the sense siRNA sometimes corresponds to (N19) TT or N21 (position 3 to 23 of the 23-nt motif), respectively. In the latter case, the 3' end of the sense siRNA often is converted to TT. The rationale for this sequence conversion is to generate a symmetric duplex with respect to the sequence composition of the sense and antisense 3' overhangs. The antisense siRNA is synthesized as the complement to position 1 to 21 of the 23-nt motif. Because position 1 of the 23-nt motif is not recognized sequence-specifically by the antisense siRNA, the 3'-most nucleotide residue of the antisense siRNA can be chosen deliberately. However, the penultimate nucleotide of the antisense siRNA (complementary to position 2 of the 23-nt motif) often is complementary to the targeted sequence. For simplifying chemical synthesis, TT often is

utilized. siRNAs corresponding to the target motif NAR(N17)YNN, where R is purine (A,G) and Y is pyrimidine (C,U), often are selected. Respective 21 nucleotide sense and antisense siRNAs often begin with a purine nucleotide and can also be expressed from pol III expression vectors without a change in targeting site. Expression of RNAs from pol III promoters often is efficient when the first transcribed nucleotide is a purine.

[0120] The sequence of the siRNA can correspond to the full length target gene, or a subsequence thereof. Often, the siRNA is about 15 to about 50 nucleotides in length (e.g., each complementary sequence of the double stranded siRNA is 15-50 nucleotides in length, and the double stranded siRNA is about 15-50 base pairs in length, sometimes about 20-30 nucleotides in length or about 20-25 nucleotides in length, e.g., 20, 21, 22, 23, 24, 25, 26, 27, 28, 29, or 30 nucleotides in length. The siRNA sometimes is about 21 nucleotides in length. Methods of using siRNA are well known in the art, and specific siRNA molecules may be purchased from a number of companies including Dharmacon Research, Inc.

[0121] Antisense, ribozyme, RNAi and siRNA nucleic acids can be altered to form modified nucleic acid molecules. The nucleic acids can be altered at base moieties, sugar moieties or phosphate backbone moieties to improve stability, hybridization, or solubility of the molecule. For example, the deoxyribose phosphate backbone of nucleic acid molecules can be modified to generate peptide nucleic acids (see Hyrup et al., *Bioorganic & Medicinal Chemistry* 4 (1): 5-23 (1996)). As used herein, the terms "peptide nucleic acid" or "PNA" refers to a nucleic acid mimic such as a DNA mimic, in which the deoxyribose phosphate backbone is replaced by a pseudopeptide backbone and only the four natural nucleobases are retained. The neutral backbone of a PNA can allow for specific hybridization to DNA and RNA under conditions of low ionic strength. Synthesis of PNA oligomers can be performed using standard solid phase peptide synthesis protocols as described, for example, in Hyrup et al., (1996) *supra* and Perry-O'Keefe et al., *Proc. Natl. Acad. Sci.* 93: 14670-675 (1996).

[0122] PNA nucleic acids can be used in prognostic, diagnostic, and therapeutic applications. For example, PNAs can be used as antisense or antigenic agents for sequence-specific modulation of gene expression by, for example, inducing transcription or translation arrest or inhibiting replication. PNA nucleic acid molecules can also be used in the analysis of single base pair mutations in a gene, (e.g., by PNA-directed PCR clamping); as "artificial restriction enzymes" when used in combination with other enzymes, (e.g., S1 nucleases (Hyrup (1996) *supra*)); or as probes or primers for DNA sequencing or hybridization (Hyrup et al., (1996) *supra*; Perry-O'Keefe *supra*).

[0123] In other embodiments, oligonucleotides may include other appended groups such as peptides (e.g., for targeting host cell receptors *in vivo*), or agents facilitating transport across cell membranes (see e.g., Letsinger et al., *Proc. Natl. Acad. Sci. USA* 86: 6553-6556 (1989); Lemaitre et al., *Proc. Natl. Acad. Sci. USA* 84: 648-652 (1987); PCT Publication No. W088/09810) or the blood-brain barrier (see, e.g., PCT Publication No. W089/10134). In addition, oligonucleotides can be modified with hybridization-

triggered cleavage agents (See, e.g., Krol et al., Bio-Techniques 6: 958-976 (1988)) or intercalating agents. (See, e.g., Zon, Pharm. Res. 5: 539-549 (1988)). To this end, the oligonucleotide may be conjugated to another molecule, (e.g., a peptide, hybridization triggered cross-linking agent, transport agent, or hybridization-triggered cleavage agent).

[0124] Also included herein are molecular beacon oligonucleotide primer and probe molecules having one or more regions complementary to a *TM7SF3* nucleotide sequence, or a substantially identical sequence thereof, two complementary regions one having a fluorophore and one a quencher such that the molecular beacon is useful for quantifying the presence of the nucleic acid in a sample. Molecular beacon nucleic acids are described, for example, in Lizardi et al., U.S. Patent No. 5,854,033; Nazarenko et al., U.S. Patent No. 5,866,336, and Livak et al., U.S. Patent 5,876,930.

Antibodies

[0125] The term "antibody" as used herein refers to an immunoglobulin molecule or immunologically active portion thereof, i.e., an antigen-binding portion. Examples of immunologically active portions of immunoglobulin molecules include F(ab) and F(ab')₂ fragments which can be generated by treating the antibody with an enzyme such as pepsin. An antibody sometimes is a polyclonal, monoclonal, recombinant (e.g., a chimeric or humanized), fully human, non-human (e.g., murine), or a single chain antibody. An antibody may have effector function and can fix complement, and is sometimes coupled to a toxin or imaging agent.

[0126] A full-length polypeptide or antigenic peptide fragment encoded by a nucleotide sequence referenced herein can be used as an immunogen or can be used to identify antibodies made with other immunogens, e.g., cells, membrane preparations, and the like. An antigenic peptide often includes at least 8 amino acid residues of the amino acid sequences encoded by a nucleotide sequence referenced herein, or substantially identical sequence thereof, and encompasses an epitope. Antigenic peptides sometimes include 10 or more amino acids, 15 or more amino acids, 20 or more amino acids, or 30 or more amino acids. Hydrophilic and hydrophobic fragments of polypeptides sometimes are used as immunogens.

[0127] Epitopes encompassed by the antigenic peptide are regions located on the surface of the polypeptide (e.g., hydrophilic regions) as well as regions with high antigenicity. For example, an Emini surface probability analysis of the human polypeptide sequence can be used to indicate the regions that have a particularly high probability of being localized to the surface of the polypeptide and are thus likely to constitute surface residues useful for targeting antibody production. The antibody may bind an epitope on any domain or region on polypeptides described herein.

[0128] Also, chimeric, humanized, and completely human antibodies are useful for applications which include repeated administration to subjects. Chimeric and humanized monoclonal antibodies,

comprising both human and non-human portions, can be made using standard recombinant DNA techniques. Such chimeric and humanized monoclonal antibodies can be produced by recombinant DNA techniques known in the art, for example using methods described in Robinson et al International Application No. PCT/US86/02269; Akira, et al European Patent Application 184,187; Taniguchi, M., European Patent Application 171,496; Morrison et al European Patent Application 173,494; Neuberger et al PCT International Publication No. WO 86/01533; Cabilly et al U.S. Patent No. 4,816,567; Cabilly et al European Patent Application 125,023; Better et al., Science 240: 1041-1043 (1988); Liu et al., Proc. Natl. Acad. Sci. USA 84: 3439-3443 (1987); Liu et al., J. Immunol. 139: 3521-3526 (1987); Sun et al., Proc. Natl. Acad. Sci. USA 84: 214-218 (1987); Nishimura et al., Canc. Res. 47: 999-1005 (1987); Wood et al., Nature 314: 446-449 (1985); and Shaw et al., J. Natl. Cancer Inst. 80: 1553-1559 (1988); Morrison, S. L., Science 229: 1202-1207 (1985); Oi et al., BioTechniques 4: 214 (1986); Winter U.S. Patent 5,225,539; Jones et al., Nature 321: 552-525 (1986); Verhoeyan et al., Science 239: 1534; and Beidler et al., J. Immunol. 141: 4053-4060 (1988).

[0129] Completely human antibodies are particularly desirable for therapeutic treatment of human patients. Such antibodies can be produced using transgenic mice that are incapable of expressing endogenous immunoglobulin heavy and light chains genes, but which can express human heavy and light chain genes. See, for example, Lonberg and Huszar, Int. Rev. Immunol. 13: 65-93 (1995); and U.S. Patent Nos. 5,625,126; 5,633,425; 5,569,825; 5,661,016; and 5,545,806. In addition, companies such as Abgenix, Inc. (Fremont, CA) and Medarex, Inc. (Princeton, NJ), can be engaged to provide human antibodies directed against a selected antigen using technology similar to that described above. Completely human antibodies that recognize a selected epitope also can be generated using a technique referred to as "guided selection." In this approach a selected non-human monoclonal antibody (e.g., a murine antibody) is used to guide the selection of a completely human antibody recognizing the same epitope. This technology is described for example by Jespers et al., Bio/Technology 12: 899-903 (1994).

[0130] An antibody can be a single chain antibody. A single chain antibody (scFv) can be engineered (see, e.g., Colcher et al., Ann. N Y Acad. Sci. 880: 263-80 (1999); and Reiter, Clin. Cancer Res. 2: 245-52 (1996)). Single chain antibodies can be dimerized or multimerized to generate multivalent antibodies having specificities for different epitopes of the same target polypeptide.

[0131] Antibodies also may be selected or modified so that they exhibit reduced or no ability to bind an Fc receptor. For example, an antibody may be an isotype or subtype, fragment or other mutant, which does not support binding to an Fc receptor (e.g., it has a mutagenized or deleted Fc receptor binding region).

[0132] Also, an antibody (or fragment thereof) may be conjugated to a therapeutic moiety such as a cytotoxin, a therapeutic agent or a radioactive metal ion. A cytotoxin or cytotoxic agent includes any agent that is detrimental to cells. Examples include taxol, cytochalasin B, gramicidin D, ethidium

bromide, emetine, mitomycin, etoposide, tenoposide, vincristine, vinblastine, colchicin, doxorubicin, daunorubicin, dihydroxy anthracin dione, mitoxantrone, mithramycin, actinomycin D, 1 dehydrotestosterone, glucocorticoids, procaine, tetracaine, lidocaine, propranolol, and puromycin and analogs or homologs thereof. Therapeutic agents include, but are not limited to, antimetabolites (e.g., methotrexate, 6-mercaptopurine, 6-thioguanine, cytarabine, 5-fluorouracil decarbazine), alkylating agents (e.g., mechlorethamine, thiotepla chlorambucil, melphalan, carmustine (BCNU) and lomustine (CCNU), cyclophosphamide, busulfan, dibromomannitol, streptozotocin, mitomycin C, and cis-dichlorodiamine platinum (II) (DDP) cisplatin), anthracyclines (e.g., daunorubicin (formerly daunomycin) and doxorubicin), antibiotics (e.g., dactinomycin (formerly actinomycin), bleomycin, mithramycin, and anthramycin (AMC)), and anti-mitotic agents (e.g., vincristine and vinblastine).

[0133] Antibody conjugates can be used for modifying a given biological response. For example, the drug moiety may be a protein or polypeptide possessing a desired biological activity. Such proteins may include, for example, a toxin such as abrin, ricin A, pseudomonas exotoxin, or diphtheria toxin; a polypeptide such as tumor necrosis factor, gamma-interferon, alpha-interferon, nerve growth factor, platelet derived growth factor, tissue plasminogen activator; or, biological response modifiers such as, for example, lymphokines, interleukin-1 ("IL-1"), interleukin-2 ("IL-2"), interleukin-6 ("IL-6"), granulocyte macrophage colony stimulating factor ("GM-CSF"), granulocyte colony stimulating factor ("G-CSF"), or other growth factors. Also, an antibody can be conjugated to a second antibody to form an antibody heteroconjugate as described by Segal in U.S. Patent No. 4,676,980, for example.

[0134] An antibody (e.g., monoclonal antibody) can be used to isolate target polypeptides by standard techniques, such as affinity chromatography or immunoprecipitation. Moreover, an antibody can be used to detect a target polypeptide (e.g., in a cellular lysate or cell supernatant) in order to evaluate the abundance and pattern of expression of the polypeptide. Antibodies can be used diagnostically to monitor polypeptide levels in tissue as part of a clinical testing procedure, e.g., to determine the efficacy of a given treatment regimen. Detection can be facilitated by coupling (i.e., physically linking) the antibody to a detectable substance (i.e., antibody labeling). Examples of detectable substances include various enzymes, prosthetic groups, fluorescent materials, luminescent materials, bioluminescent materials, and radioactive materials. Examples of suitable enzymes include horseradish peroxidase, alkaline phosphatase, β -galactosidase, or acetylcholinesterase; examples of suitable prosthetic group complexes include streptavidin/biotin and avidin/biotin; examples of suitable fluorescent materials include umbelliferone, fluorescein, fluorescein isothiocyanate, rhodamine, dichlorotriazinylamine fluorescein, dansyl chloride or phycoerythrin; an example of a luminescent material includes luminol; examples of bioluminescent materials include luciferase, luciferin, and aequorin, and examples of suitable radioactive material include ^{125}I , ^{131}I , ^{35}S or ^3H . Also, an antibody can be utilized as a test

molecule for determining whether it can treat osteoarthritis, and as a therapeutic for administration to a subject for treating osteoarthritis.

[0135] An antibody can be made by immunizing with a purified antigen, or a fragment thereof, e.g., a fragment described herein, a membrane associated antigen, tissues, e.g., crude tissue preparations, whole cells, preferably living cells, lysed cells, or cell fractions.

[0136] Included herein are antibodies which bind only a native polypeptide, only denatured or otherwise non-native polypeptide, or which bind both, as well as those having linear or conformational epitopes. Conformational epitopes sometimes can be identified by selecting antibodies that bind to native but not denatured polypeptide. Also featured are antibodies that specifically bind to a polypeptide variant associated with osteoarthritis.

Methods for Identifying Candidate Therapeutics for Treating Osteoarthritis

[0137] Current therapies for the treatment of osteoarthritis have limited efficacy, limited tolerability and significant mechanism-based side effects, and few of the available therapies adequately address underlying defects. Current therapeutic approaches were largely developed in the absence of defined molecular targets or even a solid understanding of disease pathogenesis. Therefore, provided are methods of identifying candidate therapeutics that target biochemical pathways related to the development of osteoarthritis.

[0138] Thus, featured herein are methods for identifying a candidate therapeutic for treating osteoarthritis. The methods comprise contacting a test molecule with a target molecule in a system. A "target molecule" as used herein refers to a *TM7SF3* nucleic acid, a substantially identical nucleic acid thereof, or a fragment thereof, and an encoded polypeptide of the foregoing. The methods also comprise determining the presence or absence of an interaction between the test molecule and the target molecule, where the presence of an interaction between the test molecule and the nucleic acid or polypeptide identifies the test molecule as a candidate osteoarthritis therapeutic. The interaction between the test molecule and the target molecule may be quantified.

[0139] Test molecules and candidate therapeutics include, but are not limited to, compounds, antisense nucleic acids, siRNA molecules, ribozymes, polypeptides or proteins encoded by a *TM7SF3* nucleotide sequence, or a substantially identical sequence or fragment thereof, and immunotherapeutics (e.g., antibodies and HLA-presented polypeptide fragments). A test molecule or candidate therapeutic may act as a modulator of target molecule concentration or target molecule function in a system. A "modulator" may agonize (i.e., up-regulates) or antagonize (i.e., down-regulates) a target molecule concentration partially or completely in a system by affecting such cellular functions as DNA replication and/or DNA processing (e.g., DNA methylation or DNA repair), RNA transcription and/or RNA processing (e.g., removal of intronic sequences and/or translocation of spliced mRNA from the nucleus),

polypeptide production (e.g., translation of the polypeptide from mRNA), and/or polypeptide post-translational modification (e.g., glycosylation, phosphorylation, and proteolysis of pro-polypeptides). A modulator may also agonize or antagonize a biological function of a target molecule partially or completely, where the function may include adopting a certain structural conformation, interacting with one or more binding partners, ligand binding, catalysis (e.g., phosphorylation, dephosphorylation, hydrolysis, methylation, and isomerization), and an effect upon a cellular event (e.g., effecting progression of osteoarthritis).

[0140] As used herein, the term “system” refers to a cell free *in vitro* environment and a cell-based environment such as a collection of cells, a tissue, an organ, or an organism. A system is “contacted” with a test molecule in a variety of manners, including adding molecules in solution and allowing them to interact with one another by diffusion, cell injection, and any administration routes in an animal. As used herein, the term “interaction” refers to an effect of a test molecule on test molecule, where the effect sometimes is binding between the test molecule and the target molecule, and sometimes is an observable change in cells, tissue, or organism.

[0141] There are many standard methods for detecting the presence or absence of interaction between a test molecule and a target molecule. For example, titrametric, acidimetric, radiometric, NMR, monolayer, polarographic, spectrophotometric, fluorescent, and ESR assays probative of a target molecule interaction may be utilized.

[0142] Test molecule/target molecule interactions can be detected and/or quantified using assays known in the art. For example, an interaction can be determined by labeling the test molecule and/or the target molecule, where the label is covalently or non-covalently attached to the test molecule or target molecule. The label is sometimes a radioactive molecule such as ¹²⁵I, ¹³¹I, ³⁵S or ³H, which can be detected by direct counting of radioemission or by scintillation counting. Also, enzymatic labels such as horseradish peroxidase, alkaline phosphatase, or luciferase may be utilized where the enzymatic label can be detected by determining conversion of an appropriate substrate to product. In addition, presence or absence of an interaction can be determined without labeling. For example, a microphysiometer (e.g., Cytosensor) is an analytical instrument that measures the rate at which a cell acidifies its environment using a light-addressable potentiometric sensor (LAPS). Changes in this acidification rate can be used as an indication of an interaction between a test molecule and target molecule (McConnell, H. M. *et al.*, *Science* 257: 1906-1912 (1992)).

[0143] In cell-based systems, cells typically include a *TM7SF3* nucleic acid, an encoded polypeptide, or substantially identical nucleic acid or polypeptide thereof, and are often of mammalian origin, although the cell can be of any origin. Whole cells, cell homogenates, and cell fractions (e.g., cell membrane fractions) can be subjected to analysis. Where interactions between a test molecule with a target polypeptide are monitored, soluble and/or membrane bound forms of the polypeptide may be

utilized. Where membrane-bound forms of the polypeptide are used, it may be desirable to utilize a solubilizing agent. Examples of such solubilizing agents include non-ionic detergents such as n-octylglucoside, n-dodecylglucoside, n-dodecylmaltoside, octanoyl-N-methylglucamide, decanoyl-N-methylglucamide, Triton[®] X-100, Triton[®] X-114, Thesit[®], Isotridecypoly(ethylene glycol ether)_n, 3-[(3-cholamidopropyl)dimethylamminio]-1-propane sulfonate (CHAPS), 3-[(3-cholamidopropyl)dimethylamminio]-2-hydroxy-1-propane sulfonate (CHAPSO), or N-dodecyl-N,N-dimethyl-3-ammonio-1-propane sulfonate.

[0144] An interaction between a test molecule and target molecule also can be detected by monitoring fluorescence energy transfer (FET) (*see, e.g.*, Lakowicz *et al.*, U.S. Patent No. 5,631,169; Stavrianopoulos *et al.* U.S. Patent No. 4,868,103). A fluorophore label on a first, "donor" molecule is selected such that its emitted fluorescent energy will be absorbed by a fluorescent label on a second, "acceptor" molecule, which in turn is able to fluoresce due to the absorbed energy. Alternately, the "donor" polypeptide molecule may simply utilize the natural fluorescent energy of tryptophan residues. Labels are chosen that emit different wavelengths of light, such that the "acceptor" molecule label may be differentiated from that of the "donor". Since the efficiency of energy transfer between the labels is related to the distance separating the molecules, the spatial relationship between the molecules can be assessed. In a situation in which binding occurs between the molecules, the fluorescent emission of the "acceptor" molecule label in the assay should be maximal. An FET binding event can be conveniently measured through standard fluorometric detection means well known in the art (*e.g.*, using a fluorimeter).

[0145] In another embodiment, determining the presence or absence of an interaction between a test molecule and a target molecule can be effected by monitoring surface plasmon resonance (*see, e.g.*, Sjolander & Urbanicz, *Anal. Chem.* 63: 2338-2345 (1991) and Szabo *et al.*, *Curr. Opin. Struct. Biol.* 5: 699-705 (1995)). "Surface plasmon resonance" or "biomolecular interaction analysis (BIA)" can be utilized to detect biospecific interactions in real time, without labeling any of the interactants (*e.g.*, BIAcore). Changes in the mass at the binding surface (indicative of a binding event) result in alterations of the refractive index of light near the surface (the optical phenomenon of surface plasmon resonance (SPR)), resulting in a detectable signal which can be used as an indication of real-time reactions between biological molecules.

[0146] In another embodiment, the target molecule or test molecules are anchored to a solid phase, facilitating the detection of target molecule/test molecule complexes and separation of the complexes from free, uncomplexed molecules. The target molecule or test molecule is immobilized to the solid support. In an embodiment, the target molecule is anchored to a solid surface, and the test molecule, which is not anchored, can be labeled, either directly or indirectly, with detectable labels discussed herein.

[0147] It may be desirable to immobilize a target molecule, an anti-target molecule antibody, and/or test molecules to facilitate separation of target molecule/test molecule complexes from uncomplexed forms, as well as to accommodate automation of the assay. The attachment between a test molecule and/or target molecule and the solid support may be covalent or non-covalent (*see, e.g.*, U.S. Patent No. 6,022,688 for non-covalent attachments). The solid support may be one or more surfaces of the system, such as one or more surfaces in each well of a microtiter plate, a surface of a silicon wafer, a surface of a bead (*see, e.g.*, Lam, *Nature* 354: 82-84 (1991)) that is optionally linked to another solid support, or a channel in a microfluidic device, for example. Types of solid supports, linker molecules for covalent and non-covalent attachments to solid supports, and methods for immobilizing nucleic acids and other molecules to solid supports are well known (*see, e.g.*, U.S. Patent Nos. 6,261,776; 5,900,481; 6,133,436; and 6,022,688; and WIPO publication WO 01/18234).

[0148] In an embodiment, target molecule may be immobilized to surfaces via biotin and streptavidin. For example, biotinylated target polypeptide can be prepared from biotin-NHS (N-hydroxy-succinimide) using techniques known in the art (*e.g.*, biotinylation kit, Pierce Chemicals, Rockford, IL), and immobilized in the wells of streptavidin-coated 96 well plates (Pierce Chemical). In another embodiment, a target polypeptide can be prepared as a fusion polypeptide. For example, glutathione-S-transferase/target polypeptide fusion can be adsorbed onto glutathione sepharose beads (Sigma Chemical, St. Louis, MO) or glutathione derivitized microtiter plates, which are then combined with a test molecule under conditions conducive to complex formation (*e.g.*, at physiological conditions for salt and pH). Following incubation, the beads or microtiter plate wells are washed to remove any unbound components, or the matrix is immobilized in the case of beads, and complex formation is determined directly or indirectly as described above. Alternatively, the complexes can be dissociated from the matrix, and the level of target molecule binding or activity is determined using standard techniques.

[0149] In an embodiment, the non-immobilized component is added to the coated surface containing the anchored component. After the reaction is complete, unreacted components are removed (*e.g.*, by washing) under conditions such that a significant percentage of complexes formed will remain immobilized to the solid surface. The detection of complexes anchored on the solid surface can be accomplished in a number of manners. Where the previously non-immobilized component is pre-labeled, the detection of label immobilized on the surface indicates that complexes were formed. Where the previously non-immobilized component is not pre-labeled, an indirect label can be used to detect complexes anchored on the surface, *e.g.*, by adding a labeled antibody specific for the immobilized component, where the antibody, in turn, can be directly labeled or indirectly labeled with, *e.g.*, a labeled anti-Ig antibody.

[0150] In another embodiment, an assay is performed utilizing antibodies that specifically bind target molecule or test molecule but do not interfere with binding of the target molecule to the test

molecule. Such antibodies can be derivitized to a solid support, and unbound target molecule may be immobilized by antibody conjugation. Methods for detecting such complexes, in addition to those described above for the GST-immobilized complexes, include immunodetection of complexes using antibodies reactive with the target molecule, as well as enzyme-linked assays which rely on detecting an enzymatic activity associated with the target molecule.

[0151] Cell free assays also can be conducted in a liquid phase. In such an assay, reaction products are separated from unreacted components, by any of a number of standard techniques, including but not limited to: differential centrifugation (*see, e.g., Rivas, G., and Minton, Trends Biochem Sci Aug; 18(8): 284-7 (1993)*); chromatography (gel filtration chromatography, ion-exchange chromatography); electrophoresis (*see, e.g., Ausubel et al., eds. Current Protocols in Molecular Biology, J. Wiley: New York (1999)*); and immunoprecipitation (*see, e.g., Ausubel et al., eds., supra*). Media and chromatographic techniques are known to one skilled in the art (*see, e.g., Heegaard, J Mol. Recognit. Winter; 11(1-6): 141-8 (1998)*; Hage & Tweed, *J. Chromatogr. B Biomed. Sci. Appl. Oct 10; 699 (1-2): 499-525 (1997)*). Further, fluorescence energy transfer may also be conveniently utilized, as described herein, to detect binding without further purification of the complex from solution.

[0152] In another embodiment, modulators of target molecule expression are identified. For example, a cell or cell free mixture is contacted with a candidate compound and the expression of target mRNA or target polypeptide is evaluated relative to the level of expression of target mRNA or target polypeptide in the absence of the candidate compound. When expression of target mRNA or target polypeptide is greater in the presence of the candidate compound than in its absence, the candidate compound is identified as an agonist of target mRNA or target polypeptide expression. Alternatively, when expression of target mRNA or target polypeptide is less (*e.g., less with statistical significance*) in the presence of the candidate compound than in its absence, the candidate compound is identified as an antagonist or inhibitor of target mRNA or target polypeptide expression. The level of target mRNA or target polypeptide expression can be determined by methods described herein.

[0153] In another embodiment, binding partners that interact with a target molecule are detected. The target molecules can interact with one or more cellular or extracellular macromolecules, such as polypeptides *in vivo*, and these interacting molecules are referred to herein as "binding partners." Binding partners can agonize or antagonize target molecule biological activity. Also, test molecules that agonize or antagonize interactions between target molecules and binding partners can be useful as therapeutic molecules as they can up-regulate or down-regulated target molecule activity *in vivo* and thereby treat osteoarthritis.

[0154] Binding partners of target molecules can be identified by methods known in the art. For example, binding partners may be identified by lysing cells and analyzing cell lysates by electrophoretic techniques. Alternatively, a two-hybrid assay or three-hybrid assay can be utilized (*see, e.g., U.S. Patent*

No. 5,283,317; Zervos *et al.*, *Cell* 72:223-232 (1993); Madura *et al.*, *J. Biol. Chem.* 268: 12046-12054 (1993); Bartel *et al.*, *Biotechniques* 14: 920-924 (1993); Iwabuchi *et al.*, *Oncogene* 8: 1693-1696 (1993); and Brent WO94/10300). A two-hybrid system is based on the modular nature of most transcription factors, which consist of separable DNA-binding and activation domains. The assay often utilizes two different DNA constructs. In one construct, a *TM7SF3* nucleic acid (sometimes referred to as the "bait") is fused to a gene encoding the DNA binding domain of a known transcription factor (e.g., GAL-4). In another construct, a DNA sequence from a library of DNA sequences that encodes a potential binding partner (sometimes referred to as the "prey") is fused to a gene that encodes an activation domain of the known transcription factor. Sometimes, a *TM7SF3* nucleic acid can be fused to the activation domain. If the "bait" and the "prey" molecules interact *in vivo*, the DNA-binding and activation domains of the transcription factor are brought into close proximity. This proximity allows transcription of a reporter gene (e.g., LacZ) which is operably linked to a transcriptional regulatory site responsive to the transcription factor. Expression of the reporter gene can be detected and cell colonies containing the functional transcription factor can be isolated and used to identify the potential binding partner.

[0155] In an embodiment for identifying test molecules that antagonize or agonize complex formation between target molecules and binding partners, a reaction mixture containing the target molecule and the binding partner is prepared, under conditions and for a time sufficient to allow complex formation. The reaction mixture often is provided in the presence or absence of the test molecule. The test molecule can be included initially in the reaction mixture, or can be added at a time subsequent to the addition of the target molecule and its binding partner. Control reaction mixtures are incubated without the test molecule or with a placebo. Formation of any complexes between the target molecule and the binding partner then is detected. Decreased formation of a complex in the reaction mixture containing test molecule as compared to in a control reaction mixture indicates that the molecule antagonizes target molecule/binding partner complex formation. Alternatively, increased formation of a complex in the reaction mixture containing test molecule as compared to in a control reaction mixture indicates that the molecule agonizes target molecule/binding partner complex formation. In another embodiment, complex formation of target molecule/binding partner can be compared to complex formation of mutant target molecule/binding partner (e.g., amino acid modifications in a target polypeptide). Such a comparison can be important in those cases where it is desirable to identify test molecules that modulate interactions of mutant but not non-mutated target gene products.

[0156] The assays can be conducted in a heterogeneous or homogeneous format. In heterogeneous assays, target molecule and/or the binding partner are immobilized to a solid phase, and complexes are detected on the solid phase at the end of the reaction. In homogeneous assays, the entire reaction is carried out in a liquid phase. In either approach, the order of addition of reactants can be varied to obtain different information about the molecules being tested. For example, test compounds that agonize target

molecule/binding partner interactions can be identified by conducting the reaction in the presence of the test molecule in a competition format. Alternatively, test molecules that agonize preformed complexes, *e.g.*, molecules with higher binding constants that displace one of the components from the complex, can be tested by adding the test compound to the reaction mixture after complexes have been formed.

[0157] In a heterogeneous assay embodiment, the target molecule or the binding partner is anchored onto a solid surface (*e.g.*, a microtiter plate), while the non-anchored species is labeled, either directly or indirectly. The anchored molecule can be immobilized by non-covalent or covalent attachments. Alternatively, an immobilized antibody specific for the molecule to be anchored can be used to anchor the molecule to the solid surface. The partner of the immobilized species is exposed to the coated surface with or without the test molecule. After the reaction is complete, unreacted components are removed (*e.g.*, by washing) such that a significant portion of any complexes formed will remain immobilized on the solid surface. Where the non-immobilized species is pre-labeled, the detection of label immobilized on the surface is indicative of complex. Where the non-immobilized species is not pre-labeled, an indirect label can be used to detect complexes anchored to the surface; *e.g.*, by using a labeled antibody specific for the initially non-immobilized species. Depending upon the order of addition of reaction components, test compounds that inhibit complex formation or that disrupt preformed complexes can be detected.

[0158] In another embodiment, the reaction can be conducted in a liquid phase in the presence or absence of test molecule, where the reaction products are separated from unreacted components, and the complexes are detected (*e.g.*, using an immobilized antibody specific for one of the binding components to anchor any complexes formed in solution, and a labeled antibody specific for the other partner to detect anchored complexes). Again, depending upon the order of addition of reactants to the liquid phase, test compounds that inhibit complex or that disrupt preformed complexes can be identified.

[0159] In an alternate embodiment, a homogeneous assay can be utilized. For example, a preformed complex of the target gene product and the interactive cellular or extracellular binding partner product is prepared. One or both of the target molecule or binding partner is labeled, and the signal generated by the label(s) is quenched upon complex formation (*e.g.*, U.S. Patent No. 4,109,496 that utilizes this approach for immunoassays). Addition of a test molecule that competes with and displaces one of the species from the preformed complex will result in the generation of a signal above background. In this way, test substances that disrupt target molecule/binding partner complexes can be identified.

[0160] Candidate therapeutics for treating osteoarthritis are identified from a group of test molecules that interact with a target molecule. Test molecules are normally ranked according to the degree to which they modulate (*e.g.*, agonize or antagonize) a function associated with the target molecule (*e.g.*, DNA replication and/or processing, RNA transcription and/or processing, polypeptide production and/or processing, and/or biological function/activity), and then top ranking modulators are selected. Also,

pharmacogenomic information described herein can determine the rank of a modulator. The top 10% of ranked test molecules often are selected for further testing as candidate therapeutics, and sometimes the top 15%, 20%, or 25% of ranked test molecules are selected for further testing as candidate therapeutics. Candidate therapeutics typically are formulated for administration to a subject.

Therapeutic Formulations

[0161] Formulations and pharmaceutical compositions typically include in combination with a pharmaceutically acceptable carrier one or more target molecule modulators. The modulator often is a test molecule identified as having an interaction with a target molecule by a screening method described above. The modulator may be a compound, an antisense nucleic acid, a ribozyme, an antibody, or a binding partner. Also, formulations may comprise a target polypeptide or fragment thereof in combination with a pharmaceutically acceptable carrier.

[0162] As used herein, the term “pharmaceutically acceptable carrier” includes solvents, dispersion media, coatings, antibacterial and antifungal agents, isotonic and absorption delaying agents, and the like, compatible with pharmaceutical administration. Supplementary active compounds can also be incorporated into the compositions. Pharmaceutical compositions can be included in a container, pack, or dispenser together with instructions for administration.

[0163] A pharmaceutical composition typically is formulated to be compatible with its intended route of administration. Examples of routes of administration include parenteral, *e.g.*, intravenous, intradermal, subcutaneous, oral (*e.g.*, inhalation), transdermal (topical), transmucosal, and rectal administration. Solutions or suspensions used for parenteral, intradermal, or subcutaneous application can include the following components: a sterile diluent such as water for injection, saline solution, fixed oils, polyethylene glycols, glycerin, propylene glycol or other synthetic solvents; antibacterial agents such as benzyl alcohol or methyl parabens; antioxidants such as ascorbic acid or sodium bisulfite; chelating agents such as ethylenediaminetetraacetic acid; buffers such as acetates, citrates or phosphates and agents for the adjustment of tonicity such as sodium chloride or dextrose. pH can be adjusted with acids or bases, such as hydrochloric acid or sodium hydroxide. The parenteral preparation can be enclosed in ampoules, disposable syringes or multiple dose vials made of glass or plastic.

[0164] Oral compositions generally include an inert diluent or an edible carrier. For the purpose of oral therapeutic administration, the active compound can be incorporated with excipients and used in the form of tablets, troches, or capsules, *e.g.*, gelatin capsules. Oral compositions can also be prepared using a fluid carrier for use as a mouthwash. Pharmaceutically compatible binding agents, and/or adjuvant materials can be included as part of the composition. The tablets, pills, capsules, troches and the like can contain any of the following ingredients, or compounds of a similar nature: a binder such as microcrystalline cellulose, gum tragacanth or gelatin; an excipient such as starch or lactose, a

disintegrating agent such as alginic acid, Primogel, or corn starch; a lubricant such as magnesium stearate or Sterotes; a glidant such as colloidal silicon dioxide; a sweetening agent such as sucrose or saccharin; or a flavoring agent such as peppermint, methyl salicylate, or orange flavoring.

[0165] Pharmaceutical compositions suitable for injectable use include sterile aqueous solutions (where water soluble) or dispersions and sterile powders for the extemporaneous preparation of sterile injectable solutions or dispersion. For intravenous administration, suitable carriers include physiological saline, bacteriostatic water, Cremophor EL™ (BASF, Parsippany, NJ) or phosphate buffered saline (PBS). In all cases, the composition must be sterile and should be fluid to the extent that easy syringability exists. It should be stable under the conditions of manufacture and storage and must be preserved against the contaminating action of microorganisms such as bacteria and fungi. The carrier can be a solvent or dispersion medium containing, for example, water, ethanol, polyol (for example, glycerol, propylene glycol, and liquid polyethylene glycol, and the like), and suitable mixtures thereof. The proper fluidity can be maintained, for example, by the use of a coating such as lecithin, by the maintenance of the required particle size in the case of dispersion and by the use of surfactants. Prevention of the action of microorganisms can be achieved by various antibacterial and antifungal agents, for example, parabens, chlorobutanol, phenol, ascorbic acid, thimerosal, and the like. In many cases, it will be preferable to include isotonic agents, for example, sugars, polyalcohols such as mannitol, sorbitol, sodium chloride in the composition. Prolonged absorption of the injectable compositions can be brought about by including in the composition an agent which delays absorption, for example, aluminum monostearate and gelatin.

[0166] Sterile injectable solutions can be prepared by incorporating the active compound in the required amount in an appropriate solvent with one or a combination of ingredients enumerated above, as required, followed by filtered sterilization. Generally, dispersions are prepared by incorporating the active compound into a sterile vehicle which contains a basic dispersion medium and the required other ingredients from those enumerated above. In the case of sterile powders for the preparation of sterile injectable solutions, the preferred methods of preparation are vacuum drying and freeze-drying which yields a powder of the active ingredient plus any additional desired ingredient from a previously sterile-filtered solution thereof.

[0167] For administration by inhalation, the compounds are delivered in the form of an aerosol spray from pressured container or dispenser which contains a suitable propellant, *e.g.*, a gas such as carbon dioxide, or a nebulizer.

[0168] Systemic administration can also be by transmucosal or transdermal means. For transmucosal or transdermal administration, penetrants appropriate to the barrier to be permeated are used in the formulation. Such penetrants are generally known in the art, and include, for example, for transmucosal administration, detergents, bile salts, and fusidic acid derivatives. Transmucosal administration can be accomplished through the use of nasal sprays or suppositories. For transdermal

administration, the active compounds are formulated into ointments, salves, gels, or creams as generally known in the art. Molecules can also be prepared in the form of suppositories (*e.g.*, with conventional suppository bases such as cocoa butter and other glycerides) or retention enemas for rectal delivery.

[0169] In one embodiment, active molecules are prepared with carriers that will protect the compound against rapid elimination from the body, such as a controlled release formulation, including implants and microencapsulated delivery systems. Biodegradable, biocompatible polymers can be used, such as ethylene vinyl acetate, polyanhydrides, polyglycolic acid, collagen, polyorthoesters, and polylactic acid. Methods for preparation of such formulations will be apparent to those skilled in the art. Materials can also be obtained commercially from Alza Corporation and Nova Pharmaceuticals, Inc. Liposomal suspensions (including liposomes targeted to infected cells with monoclonal antibodies to viral antigens) can also be used as pharmaceutically acceptable carriers. These can be prepared according to methods known to those skilled in the art, for example, as described in U.S. Patent No. 4,522,811.

[0170] It is advantageous to formulate oral or parenteral compositions in dosage unit form for ease of administration and uniformity of dosage. Dosage unit form as used herein refers to physically discrete units suited as unitary dosages for the subject to be treated; each unit containing a predetermined quantity of active compound calculated to produce the desired therapeutic effect in association with the required pharmaceutical carrier.

[0171] Toxicity and therapeutic efficacy of such compounds can be determined by standard pharmaceutical procedures in cell cultures or experimental animals, *e.g.*, for determining the LD₅₀ (the dose lethal to 50% of the population) and the ED₅₀ (the dose therapeutically effective in 50% of the population). The dose ratio between toxic and therapeutic effects is the therapeutic index and it can be expressed as the ratio LD₅₀/ED₅₀. Molecules which exhibit high therapeutic indices are preferred. While molecules that exhibit toxic side effects may be used, care should be taken to design a delivery system that targets such compounds to the site of affected tissue in order to minimize potential damage to uninfected cells and, thereby, reduce side effects.

[0172] The data obtained from the cell culture assays and animal studies can be used in formulating a range of dosage for use in humans. The dosage of such molecules lies preferably within a range of circulating concentrations that include the ED₅₀ with little or no toxicity. The dosage may vary within this range depending upon the dosage form employed and the route of administration utilized. For any molecules used in the methods described herein, the therapeutically effective dose can be estimated initially from cell culture assays. A dose may be formulated in animal models to achieve a circulating plasma concentration range that includes the IC₅₀ (*i.e.*, the concentration of the test compound which achieves a half-maximal inhibition of symptoms) as determined in cell culture. Such information can be

used to more accurately determine useful doses in humans. Levels in plasma may be measured, for example, by high performance liquid chromatography.

[0173] As defined herein, a therapeutically effective amount of protein or polypeptide (*i.e.*, an effective dosage) ranges from about 0.001 to 30 mg/kg body weight, sometimes about 0.01 to 25 mg/kg body weight, often about 0.1 to 20 mg/kg body weight, and more often about 1 to 10 mg/kg, 2 to 9 mg/kg, 3 to 8 mg/kg, 4 to 7 mg/kg, or 5 to 6 mg/kg body weight. The protein or polypeptide can be administered one time per week for between about 1 to 10 weeks, sometimes between 2 to 8 weeks, often between about 3 to 7 weeks, and more often for about 4, 5, or 6 weeks. The skilled artisan will appreciate that certain factors may influence the dosage and timing required to effectively treat a subject, including but not limited to the severity of the disease or disorder, previous treatments, the general health and/or age of the subject, and other diseases present. Moreover, treatment of a subject with a therapeutically effective amount of a protein, polypeptide, or antibody can include a single treatment or, preferably, can include a series of treatments.

[0174] With regard to polypeptide formulations, featured herein is a method for treating osteoarthritis in a subject, which comprises contacting one or more cells in the subject with a first polypeptide, where the subject comprises a second polypeptide having one or more polymorphic variations associated with cancer, and where the first polypeptide comprises fewer polymorphic variations associated with cancer than the second polypeptide. The first and second polypeptides are encoded by a nucleic acid which comprises a nucleotide sequence in SEQ ID NO: 1-2; a nucleotide sequence which encodes a polypeptide consisting of an amino acid sequence encoded by a nucleotide sequence referenced in SEQ ID NO: 1-2; a nucleotide sequence which encodes a polypeptide that is 90% or more identical to an amino acid sequence encoded by a nucleotide sequence of SEQ ID NO: 1-2 and a nucleotide sequence 90% or more identical to a nucleotide sequence in SEQ ID NO: 1-2. The subject often is a human.

[0175] For antibodies, a dosage of 0.1 mg/kg of body weight (generally 10 mg/kg to 20 mg/kg) is often utilized. If the antibody is to act in the brain, a dosage of 50 mg/kg to 100 mg/kg is often appropriate. Generally, partially human antibodies and fully human antibodies have a longer half-life within the human body than other antibodies. Accordingly, lower dosages and less frequent administration is often possible. Modifications such as lipidation can be used to stabilize antibodies and to enhance uptake and tissue penetration (*e.g.*, into the brain). A method for lipidation of antibodies is described by Cruikshank *et al.*, *J. Acquired Immune Deficiency Syndromes and Human Retrovirology* 14:193 (1997).

[0176] Antibody conjugates can be used for modifying a given biological response, the drug moiety is not to be construed as limited to classical chemical therapeutic agents. For example, the drug moiety may be a protein or polypeptide possessing a desired biological activity. Such proteins may include, for

example, a toxin such as abrin, ricin A, pseudomonas exotoxin, or diphtheria toxin; a polypeptide such as tumor necrosis factor, alpha-interferon, beta-interferon, nerve growth factor, platelet derived growth factor, tissue plasminogen activator; or, biological response modifiers such as, for example, lymphokines, interleukin-1 ("IL-1"), interleukin-2 ("IL-2"), interleukin-6 ("IL-6"), granulocyte macrophage colony stimulating factor ("GM-CSF"), granulocyte colony stimulating factor ("G-CSF"), or other growth factors. Alternatively, an antibody can be conjugated to a second antibody to form an antibody heteroconjugate as described by Segal in U.S. Patent No. 4,676,980.

[0177] For compounds, exemplary doses include milligram or microgram amounts of the compound per kilogram of subject or sample weight, for example, about 1 microgram per kilogram to about 500 milligrams per kilogram, about 100 micrograms per kilogram to about 5 milligrams per kilogram, or about 1 microgram per kilogram to about 50 micrograms per kilogram. It is understood that appropriate doses of a small molecule depend upon the potency of the small molecule with respect to the expression or activity to be modulated. When one or more of these small molecules is to be administered to an animal (e.g., a human) in order to modulate expression or activity of a polypeptide or nucleic acid described herein, a physician, veterinarian, or researcher may, for example, prescribe a relatively low dose at first, subsequently increasing the dose until an appropriate response is obtained. In addition, it is understood that the specific dose level for any particular animal subject will depend upon a variety of factors including the activity of the specific compound employed, the age, body weight, general health, gender, and diet of the subject, the time of administration, the route of administration, the rate of excretion, any drug combination, and the degree of expression or activity to be modulated.

[0178] With regard to nucleic acid formulations, gene therapy vectors can be delivered to a subject by, for example, intravenous injection, local administration (*see, e.g.,* U.S. Patent 5,328,470) or by stereotactic injection (*see e.g., Chen et al., (1994) Proc. Natl. Acad. Sci. USA 91:3054-3057*). Pharmaceutical preparations of gene therapy vectors can include a gene therapy vector in an acceptable diluent, or can comprise a slow release matrix in which the gene delivery vehicle is imbedded. Alternatively, where the complete gene delivery vector can be produced intact from recombinant cells (e.g., retroviral vectors) the pharmaceutical preparation can include one or more cells which produce the gene delivery system. Examples of gene delivery vectors are described herein.

Therapeutic Methods

[0179] A therapeutic formulation described above can be administered to a subject in need of a therapeutic for inducing a desired biological response.. Therapeutic formulations can be administered by any of the paths described herein. With regard to both prophylactic and therapeutic methods of treatment, such treatments may be specifically tailored or modified, based on knowledge obtained from pharmacogenomic analyses described herein.

[0180] As used herein, the term “treatment” is defined as the application or administration of a therapeutic formulation to a subject, or application or administration of a therapeutic agent to an isolated tissue or cell line from a subject with the purpose to cure, heal, alleviate, relieve, alter, remedy, ameliorate, improve or affect osteoarthritis, symptoms of osteoarthritis or a predisposition towards osteoarthritis. A therapeutic formulation includes, but is not limited to, small molecules, peptides, antibodies, ribozymes and antisense oligonucleotides. Administration of a therapeutic formulation can occur prior to the manifestation of symptoms characteristic of osteoarthritis, such that osteoarthritis is prevented or delayed in its progression. The appropriate therapeutic composition can be determined based on screening assays described herein.

[0181] As discussed, successful treatment of osteoarthritis can be brought about by techniques that serve to agonize target molecule expression or function, or alternatively, antagonize target molecule expression or function. These techniques include administration of modulators that include, but are not limited to, small organic or inorganic molecules; antibodies (including, for example, polyclonal, monoclonal, humanized, anti-idiotypic, chimeric or single chain antibodies, and Fab, F(ab')₂ and Fab expression library fragments, scFV molecules, and epitope-binding fragments thereof); and peptides, phosphopeptides, or polypeptides.

[0182] Further, antisense and ribozyme molecules that inhibit expression of the target gene can also be used to reduce the level of target gene expression, thus effectively reducing the level of target gene activity. Still further, triple helix molecules can be utilized in reducing the level of target gene activity. Antisense, ribozyme and triple helix molecules are discussed above. It is possible that the use of antisense, ribozyme, and/or triple helix molecules to reduce or inhibit mutant gene expression can also reduce or inhibit the transcription (triple helix) and/or translation (antisense, ribozyme) of mRNA produced by normal target gene alleles, such that the concentration of normal target gene product present can be lower than is necessary for a normal phenotype. In such cases, nucleic acid molecules that encode and express target gene polypeptides exhibiting normal target gene activity can be introduced into cells via gene therapy method. Alternatively, in instances in that the target gene encodes an extracellular polypeptide, it can be preferable to co-administer normal target gene polypeptide into the cell or tissue in order to maintain the requisite level of cellular or tissue target gene activity.

[0183] Another method by which nucleic acid molecules may be utilized in treating or preventing osteoarthritis is use of aptamer molecules specific for target molecules. Aptamers are nucleic acid molecules having a tertiary structure which permits them to specifically bind to ligands (*see, e.g., Osborne, et al., Curr. Opin. Chem. Biol.* 1(1): 5-9 (1997); and Patel, D. J., *Curr. Opin. Chem. Biol.* Jun; 1(1): 32-46 (1997)).

[0184] Yet another method of utilizing nucleic acid molecules for osteoarthritis treatment is gene therapy, which can also be referred to as allele therapy. Provided herein is a gene therapy method for

treating osteoarthritis in a subject, which comprises contacting one or more cells in the subject or from the subject with a nucleic acid having a first nucleotide sequence (e.g., the first nucleotide sequence is identical to or substantially identical to a nucleotide sequence of SEQ ID NO: 1-2). Genomic DNA in the subject comprises a second nucleotide sequence having one or more polymorphic variations associated with osteoarthritis (e.g., the second nucleotide sequence is identical to or substantially identical to a nucleotide sequence of SEQ ID NO: 1). The first and second nucleotide sequences typically are substantially identical to one another, and the first nucleotide sequence comprises fewer polymorphic variations associated with osteoarthritis than the second nucleotide sequence. The first nucleotide sequence may comprise a gene sequence that encodes a full-length polypeptide or a fragment thereof. The subject is often a human. Allele therapy methods often are utilized in conjunction with a method of first determining whether a subject has genomic DNA that includes polymorphic variants associated with osteoarthritis.

[0185] In another allele therapy embodiment, provided herein is a method which comprises contacting one or more cells in the subject or from the subject with a polypeptide encoded by a nucleic acid having a first nucleotide sequence (e.g., the first nucleotide sequence is identical to or substantially identical to the nucleotide sequence of SEQ ID NO: 1-2). Genomic DNA in the subject comprises a second nucleotide sequence having one or more polymorphic variations associated with osteoarthritis (e.g., the second nucleotide sequence is identical to or substantially identical to a nucleotide sequence of SEQ ID NO: 1). The first and second nucleotide sequences typically are substantially identical to one another, and the first nucleotide sequence comprises fewer polymorphic variations associated with osteoarthritis than the second nucleotide sequence. The first nucleotide sequence may comprise a gene sequence that encodes a full-length polypeptide or a fragment thereof. The subject is often a human.

[0186] For antibody-based therapies, antibodies can be generated that are both specific for target molecules and that reduce target molecule activity. Such antibodies may be administered in instances where antagonizing a target molecule function is appropriate for the treatment of osteoarthritis.

[0187] In circumstances where stimulating antibody production in an animal or a human subject by injection with a target molecule is harmful to the subject, it is possible to generate an immune response against the target molecule by use of anti-idiotypic antibodies (*see, e.g., Herlyn, Ann. Med.*; 31(1): 66-78 (1999); and Bhattacharya-Chatterjee & Foon, *Cancer Treat. Res.*; 94: 51-68 (1998)). Introducing an anti-idiotypic antibody to a mammal or human subject often stimulates production of anti-anti-idiotypic antibodies, which typically are specific to the target molecule. Vaccines directed to osteoarthritis also may be generated in this fashion.

[0188] In instances where the target molecule is intracellular and whole antibodies are used, internalizing antibodies may be preferred. Lipofectin or liposomes can be used to deliver the antibody or a fragment of the Fab region that binds to the target antigen into cells. Where fragments of the antibody

are used, the smallest inhibitory fragment that binds to the target antigen is preferred. For example, peptides having an amino acid sequence corresponding to the Fv region of the antibody can be used. Alternatively, single chain neutralizing antibodies that bind to intracellular target antigens can also be administered. Such single chain antibodies can be administered, for example, by expressing nucleotide sequences encoding single-chain antibodies within the target cell population (*see, e.g., Marasco et al., Proc. Natl. Acad. Sci. USA 90: 7889-7893 (1993)*).

[0189] Modulators can be administered to a patient at therapeutically effective doses to treat osteoarthritis. A therapeutically effective dose refers to an amount of the modulator sufficient to result in amelioration of symptoms of osteoarthritis. Toxicity and therapeutic efficacy of modulators can be determined by standard pharmaceutical procedures in cell cultures or experimental animals, *e.g.,* for determining the LD₅₀ (the dose lethal to 50% of the population) and the ED₅₀ (the dose therapeutically effective in 50% of the population). The dose ratio between toxic and therapeutic effects is the therapeutic index and it can be expressed as the ratio LD₅₀/ED₅₀. Modulators that exhibit large therapeutic indices are preferred. While modulators that exhibit toxic side effects can be used, care should be taken to design a delivery system that targets such molecules to the site of affected tissue in order to minimize potential damage to uninfected cells, thereby reducing side effects.

[0190] Data obtained from cell culture assays and animal studies can be used in formulating a range of dosages for use in humans. The dosage of such compounds lies preferably within a range of circulating concentrations that include the ED₅₀ with little or no toxicity. The dosage can vary within this range depending upon the dosage form employed and the route of administration utilized. For any compound used in the methods described herein, the therapeutically effective dose can be estimated initially from cell culture assays. A dose can be formulated in animal models to achieve a circulating plasma concentration range that includes the IC₅₀ (*i.e.,* the concentration of the test compound that achieves a half-maximal inhibition of symptoms) as determined in cell culture. Such information can be used to more accurately determine useful doses in humans. Levels in plasma can be measured, for example, by high performance liquid chromatography.

[0191] Another example of effective dose determination for an individual is the ability to directly assay levels of "free" and "bound" compound in the serum of the test subject. Such assays may utilize antibody mimics and/or "biosensors" that have been created through molecular imprinting techniques. Molecules that modulate target molecule activity are used as a template, or "imprinting molecule", to spatially organize polymerizable monomers prior to their polymerization with catalytic reagents. The subsequent removal of the imprinted molecule leaves a polymer matrix which contains a repeated "negative image" of the compound and is able to selectively rebinding the molecule under biological assay conditions. A detailed review of this technique can be seen in Ansell *et al., Current Opinion in Biotechnology 7: 89-94 (1996)* and in Shea, *Trends in Polymer Science 2: 166-173 (1994)*. Such

“imprinted” affinity matrixes are amenable to ligand-binding assays, whereby the immobilized monoclonal antibody component is replaced by an appropriately imprinted matrix. An example of the use of such matrixes in this way can be seen in Vlatakis, *et al.*, *Nature* 361: 645-647 (1993). Through the use of isotope-labeling, the “free” concentration of compound which modulates target molecule expression or activity readily can be monitored and used in calculations of IC_{50} . Such “imprinted” affinity matrixes can also be designed to include fluorescent groups whose photon-emitting properties measurably change upon local and selective binding of target compound. These changes readily can be assayed in real time using appropriate fiberoptic devices, in turn allowing the dose in a test subject to be quickly optimized based on its individual IC_{50} . An example of such a “biosensor” is discussed in Kriz *et al.*, *Analytical Chemistry* 67: 2142-2144 (1995).

[0192] The examples set forth below are intended to illustrate but not limit the invention.

Examples

[0193] In the following studies a group of subjects was selected according to specific parameters relating to osteoarthritis (OA). Nucleic acid samples obtained from individuals in the study group were subjected to genetic analysis, which identified associations between osteoarthritis and a polymorphism in the *TM7SF3* gene on chromosome twelve. The polymorphism was genotyped again in two replication cohorts consisting of individuals selected for OA. In addition, SNPs proximal to the incident polymorphism were identified and allelotyped in OA case and control pools. Methods are described for producing *TM7SF3* polypeptide and *TM7SF3* polypeptide variants *in vitro* or *in vivo*, *TM7SF3* nucleic acids or polypeptides and variants thereof are utilized for screening test molecules for those that interact with *TM7SF3* molecules. Test molecules identified as interactors with *TM7SF3* molecules and *TM7SF3* variants are further screened *in vivo* to determine whether they treat osteoarthritis.

Example 1

Samples and Pooling Strategies

Sample Selection

[0194] Blood samples were collected from individuals diagnosed with knee osteoarthritis, which were referred to as case samples. Also, blood samples were collected from individuals not diagnosed with knee osteoarthritis as gender and age-matched controls. A database was created that listed all phenotypic trait information gathered from individuals for each case and control sample. Genomic DNA was extracted from each of the blood samples for genetic analyses.

DNA Extraction from Blood Samples

[0195] Six to ten milliliters of whole blood was transferred to a 50 ml tube containing 27 ml of red cell lysis solution (RCL). The tube was inverted until the contents were mixed. Each tube was incubated for 10 minutes at room temperature and inverted once during the incubation. The tubes were then centrifuged for 20 minutes at 3000 x g and the supernatant was carefully poured off. 100-200 μ l of residual liquid was left in the tube and was pipetted repeatedly to resuspend the pellet in the residual supernatant. White cell lysis solution (WCL) was added to the tube and pipetted repeatedly until completely mixed. While no incubation was normally required, the solution was incubated at 37°C or room temperature if cell clumps were visible after mixing until the solution was homogeneous. 2 ml of protein precipitation was added to the cell lysate. The mixtures were vortexed vigorously at high speed for 20 sec to mix the protein precipitation solution uniformly with the cell lysate, and then centrifuged for 10 minutes at 3000 x g. The supernatant containing the DNA was then poured into a clean 15 ml tube, which contained 7 ml of 100% isopropanol. The samples were mixed by inverting the tubes gently until white threads of DNA were visible. Samples were centrifuged for 3 minutes at 2000 x g and the DNA was visible as a small white pellet. The supernatant was decanted and 5 ml of 70% ethanol was added to each tube. Each tube was inverted several times to wash the DNA pellet, and then centrifuged for 1 minute at 2000 x g. The ethanol was decanted and each tube was drained on clean absorbent paper. The DNA was dried in the tube by inversion for 10 minutes, and then 1000 μ l of 1X TE was added. The size of each sample was estimated, and less TE buffer was added during the following DNA hydration step if the sample was smaller. The DNA was allowed to rehydrate overnight at room temperature, and DNA samples were stored at 2-8°C.

[0196] DNA was quantified by placing samples on a hematology mixer for at least 1 hour. DNA was serially diluted (typically 1:80, 1:160, 1:320, and 1:640 dilutions) so that it would be within the measurable range of standards. 125 μ l of diluted DNA was transferred to a clear U-bottom microtitre plate, and 125 μ l of 1X TE buffer was transferred into each well using a multichannel pipette. The DNA and 1X TE were mixed by repeated pipetting at least 15 times, and then the plates were sealed. 50 μ l of diluted DNA was added to wells A5-H12 of a black flat bottom microtitre plate. Standards were inverted six times to mix them, and then 50 μ l of 1X TE buffer was pipetted into well A1, 1000 ng/ml of standard was pipetted into well A2, 500 ng/ml of standard was pipetted into well A3, and 250 ng/ml of standard was pipetted into well A4. PicoGreen (Molecular Probes, Eugene, Oregon) was thawed and freshly diluted 1:200 according to the number of plates that were being measured. PicoGreen was vortexed and then 50 μ l was pipetted into all wells of the black plate with the diluted DNA. DNA and PicoGreen were mixed by pipetting repeatedly at least 10 times with the multichannel pipette. The plate was placed into a Fluoroskan Ascent Machine (microplate fluorometer produced by Labsystems) and the samples were

allowed to incubate for 3 minutes before the machine was run using filter pairs 485 nm excitation and 538 nm emission wavelengths. Samples having measured DNA concentrations of greater than 450 ng/ μ l were re-measured for conformation. Samples having measured DNA concentrations of 20 ng/ μ l or less were re-measured for confirmation.

Pooling Strategies – Discovery Cohort

[0197] Samples were derived from the Nottingham knee OA family study (UK) where index cases were identified through a knee replacement registry. Siblings were approached and assessed with knee x-rays and assigned status as affected or unaffected. In all 1,157 individuals were available. In order to create same-sex pools of appropriate sizes, 335 unrelated female individuals with OA from the Nottingham OA sample were selected for the case pool. The control pool was made up of unrelated female individuals from the St. Thomas twin study (England) with normal knee x-rays and without other indications of OA, regardless of anatomical location, as well as lacking family history of OA. The St. Thomas twin study consists of Caucasian, female participants from the St. Thomas' Hospital, London, adult-twin registry, which is a voluntary registry of >4,000 twin pairs ranging from 18 to 76 years of age. The female case samples and female control samples are described further in Table 1 below.

[0198] A select set of samples from each group were utilized to generate pools, and one pool was created for each group. Each individual sample in a pool was represented by an equal amount of genomic DNA. For example, where 25 ng of genomic DNA was utilized in each PCR reaction and there were 200 individuals in each pool, each individual would provide 125 pg of genomic DNA. Inclusion or exclusion of samples for a pool was based upon the following criteria: the sample was derived from an individual characterized as Caucasian; the sample was derived from an individual of British paternal and maternal descent; case samples were derived from individuals diagnosed with specific knee osteoarthritis (OA) and were recruited from an OA knee replacement clinic. Control samples were derived from individuals free of OA, family history of OA, and rheumatoid arthritis. Also, sufficient genomic DNA was extracted from each blood sample for all allelotyping and genotyping reactions performed during the study. Phenotype information from each individual was collected and included age of the individual, gender, family history of OA, general medical information (e.g., height, weight, thyroid disease, diabetes, psoriasis, hysterectomy), joint history (previous and current symptoms, joint-related operations, age at onset of symptoms, date of primary diagnosis, age of individual as of primary diagnosis and order of involvement), and knee-related findings (crepitus, restricted passive movement, bony swelling/deformity). Additional knee information included knee history, current symptoms, any major knee injury, meniscectomy, knee replacement surgery, age of surgery, and treatment history (including

hormone replace therapy (HRT)). Samples that met these criteria were added to appropriate pools based on disease status.

[0199] The selection process yielded the pools set forth in Table 1, which were used in the studies that follow:

TABLE 1

	Female case	Female control
Pool size (Number)	335	335
Pool Criteria (ex: case/control)	control	case
Mean Age (ex: years)	57.21	69.95

Example 2

Association of Polymorphic Variants with Osteoarthritis

[0200] A whole-genome screen was performed to identify particular SNPs associated with occurrence of osteoarthritis. As described in Example 1, two sets of samples were utilized, which included samples from female individuals having knee osteoarthritis (osteoarthritis cases), and samples from female individuals not having knee osteoarthritis (female controls). The initial screen of each pool was performed in an allelotyping study, in which certain samples in each group were pooled. By pooling DNA from each group, an allele frequency for each SNP in each group was calculated. These allele frequencies were then compared to one another. Particular SNPs were considered as being associated with osteoarthritis when allele frequency differences calculated between case and control pools were statistically significant. SNP disease association results obtained from the allelotyping study were then validated by genotyping each associated SNP across all samples from each pool. The results of the genotyping then were analyzed, allele frequencies for each group were calculated from the individual genotyping results, and a p-value was calculated to determine whether the case and control groups had statistically significant differences in allele frequencies for a particular SNP. When the genotyping results agreed with the original allelotyping results, the SNP disease association was considered validated at the genetic level.

SNP Panel Used for Genetic Analyses

[0201] A whole-genome SNP screen began with an initial screen of approximately 25,000 SNPs over each set of disease and control samples using a pooling approach. The pools studied in the screen are described in Example 1. The SNPs analyzed in this study were part of a set of 25,488 SNPs

confirmed as being statistically polymorphic as each is characterized as having a minor allele frequency of greater than 10%. The SNPs in the set reside in genes or in close proximity to genes, and many reside in gene exons. Specifically, SNPs in the set are located in exons, introns, and within 5,000 base-pairs upstream of a transcription start site of a gene. In addition, SNPs were selected according to the following criteria: they are located in ESTs; they are located in Locuslink or Ensembl genes; and they are located in Genomatix promoter predictions. SNPs in the set were also selected on the basis of even spacing across the genome, as depicted in Table 2.

[0202] A case-control study design using a whole genome association strategy involving approximately 28,000 single nucleotide polymorphisms (SNPs) was employed. Approximately 25,000 SNPs were evenly spaced in gene-based regions of the human genome with a median inter-marker distance of about 40,000 base pairs. Additionally, approximately 3,000 SNPs causing amino acid substitutions in genes described in the literature as candidates for various diseases were used. The case-control study samples were of female Caucasian origin (British paternal and maternal descent) 670 individuals were equally distributed in two groups: female controls and female cases. The whole genome association approach was first conducted on 2 DNA pools representing the 2 groups. Significant markers were confirmed by individual genotyping.

TABLE 2

<u>General Statistics</u>		<u>Spacing Statistics</u>	
Total # of SNPs	25,488	Median	37,058 bp
# of Exonic SNPs	>4,335 (17%)	Minimum*	1,000 bp
# SNPs with refSNP ID	20,776 (81%)	Maximum*	3,000,000 bp
Gene Coverage	>10,000	Mean	122,412 bp
Chromosome Coverage	All	Std Deviation	373,325 bp
		<i>*Excludes outliers</i>	

Allelotyping and Genotyping Results

[0203] The genetic studies summarized above and described in more detail below identified an allelic variant in the TM7SF3 gene that is associated with osteoarthritis.

Assay for Verifying, Allelotyping, and Genotyping SNPs

[0204] A MassARRAY™ system (Sequenom, Inc.) was utilized to perform SNP genotyping in a high-throughput fashion. This genotyping platform was complemented by a homogeneous, single-tube assay method (hMET™ or homogeneous MassEXTEND™ (Sequenom, Inc.)) in which two genotyping primers anneal to and amplify a genomic target surrounding a polymorphic site of interest. A third

primer (the MassEXTEND™ primer), which is complementary to the amplified target up to but not including the polymorphism, was then enzymatically extended one or a few bases through the polymorphic site and then terminated.

[0205] For each polymorphism, SpectroDESIGNER™ software (Sequenom, Inc.) was used to generate a set of PCR primers and a MassEXTEND™ primer which were used to genotype the polymorphism. Other primer design software could be used or one of ordinary skill in the art could manually design primers based on his or her knowledge of the relevant factors and considerations in designing such primers. Table 3 shows PCR primers and Table 4 shows extension primers used for analyzing polymorphisms. The initial PCR amplification reaction was performed in a 5 µl total volume containing 1X PCR buffer with 1.5 mM MgCl₂ (Qiagen), 200 µM each of dATP, dGTP, dCTP, dTTP (Gibco-BRL), 2.5 ng of genomic DNA, 0.1 units of HotStar DNA polymerase (Qiagen), and 200 nM each of forward and reverse PCR primers specific for the polymorphic region of interest.

TABLE 3: PCR Primers

SNP Reference	Forward PCR primer	Reverse PCR primer
rs1484086	ACGTTGGATGTGTCACTCTTCGGAAGTCTC	ACGTTGGATGCATGTACAGGGCATTACAG

[0206] Samples were incubated at 95°C for 15 minutes, followed by 45 cycles of 95°C for 20 seconds, 56°C for 30 seconds, and 72°C for 1 minute, finishing with a 3 minute final extension at 72°C. Following amplification, shrimp alkaline phosphatase (SAP) (0.3 units in a 2 µl volume) (Amersham Pharmacia) was added to each reaction (total reaction volume was 7 µl) to remove any residual dNTPs that were not consumed in the PCR step. Samples were incubated for 20 minutes at 37°C, followed by 5 minutes at 85°C to denature the SAP.

[0207] Once the SAP reaction was complete, a primer extension reaction was initiated by adding a polymorphism-specific MassEXTEND™ primer cocktail to each sample. Each MassEXTEND™ cocktail included a specific combination of dideoxynucleotides (ddNTPs) and deoxynucleotides (dNTPs) used to distinguish polymorphic alleles from one another. Methods for verifying, allelotyping and genotyping SNPs are disclosed, for example, in U.S. Pat. No. 6,258,538, the content of which is hereby incorporated by reference. In Table 4, ddNTPs are shown and the fourth nucleotide not shown is the dNTP.

TABLE 4: Extension Primers

SNP Reference	Extend Probe	Termination Mix
rs1484086	CTCTTCGGAAGTCTCTTCTCA	ACT

[0208] The MassEXTEND™ reaction was performed in a total volume of 9 µl, with the addition of 1X ThermoSequenase buffer, 0.576 units of ThermoSequenase (Amersham Pharmacia), 600 nM MassEXTEND™ primer, 2 mM of ddATP and/or ddCTP and/or ddGTP and/or ddTTP, and 2 mM of dATP or dCTP or dGTP or dTTP. The deoxy nucleotide (dNTP) used in the assay normally was complementary to the nucleotide at the polymorphic site in the amplicon. Samples were incubated at 94°C for 2 minutes, followed by 55 cycles of 5 seconds at 94°C, 5 seconds at 52°C, and 5 seconds at 72°C.

[0209] Following incubation, samples were desalted by adding 16 µl of water (total reaction volume was 25 µl), 3 mg of SpectroCLEAN™ sample cleaning beads (Sequenom, Inc.) and allowed to incubate for 3 minutes with rotation. Samples were then robotically dispensed using a piezoelectric dispensing device (SpectroJET™ (Sequenom, Inc.)) onto either 96-spot or 384-spot silicon chips containing a matrix that crystallized each sample (SpectroCHIP™ (Sequenom, Inc.)). Subsequently, MALDI-TOF mass spectrometry (Biflex and Autoflex MALDI-TOF mass spectrometers (Bruker Daltonics) can be used) and SpectroTYPER RT™ software (Sequenom, Inc.) were used to analyze and interpret the SNP genotype for each sample.

Genetic Analysis

[0210] Minor allelic frequencies for the polymorphisms set forth in Table A were verified as being 10% or greater using the extension assay described above in a group of samples isolated from 92 individuals originating from the state of Utah in the United States, Venezuela and France (Coriell cell repositories).

[0211] Genotyping results are shown for female pools in Table 5. In Table 5, “AF” refers to allelic frequency; and “F case” and “F control” refer to female case and female control groups, respectively.

TABLE 5: Genotyping Results

SNP Reference	AF F case	AF F control	p-value
rs1484086	T = 0.821 C = 0.179	T = 0.753 C = 0.247	0.0027

[0212] The single marker alleles set forth in Table A were considered validated, since the genotyping data agreed with the allelotyping data and each SNP significantly associated with osteoarthritis. Particularly significant associations with osteoarthritis are indicated by a calculated p-value of less than 0.05 for genotype results.

Example 3

Association of Polymorphic Variants with Osteoarthritis in Replication Cohorts

[0213] The single marker polymorphism set forth in Table A was genotyped again in two replication cohorts consisting of individuals selected for OA.

Sample Selection and Pooling Strategies – Replication Sample 1

[0214] A second case control sample (replication sample #1) was created by using 100 Caucasian female cases from Chingford, UK, and 148 unrelated female cases from the St. Thomas twin study. Cases were defined as having Kellgren-Lawrence (KL) scores of at least 2 in at least one knee x-ray. In addition, 199 male knee replacement cases from Nottingham were included. (For a cohort description, see the Nottingham description provided in Example 1). The control pool was made up of unrelated female individuals from the St. Thomas twin study (England) with normal knee x-rays and without other indications of OA, regardless of anatomical location, as well as lacking family history of OA. The St. Thomas twin study consists of Caucasian, female participants from the St. Thomas' Hospital, London, adult-twin registry, which is a voluntary registry of >4,000 twin pairs ranging from 18 to 76 years of age. The replication sample 1 cohort was used to replicate the initial results. Table 6 below summarizes the selected phenotype data collected from the case and control individuals.

TABLE 6

Phenotype	Female cases (n=248): median (range)/ (n,%)	Male cases (n=199): median (range)/ (n,%)	Female controls (n=313): mean (range)/ (n,%)
Age	59 (39- 73)	66 (45- 73)	55 (50- 72)
Height (cm)	162 (141- 178)	175 (152- 198)	162 (141- 176)
Weight (kg)	68 (51- 123)	86 (62- 127)	64 (40- 111)
Body mass index (kg/m ²)	26 (18- 44)	29 (21- 41)	24 (18- 46)
Kellgren- Lawrence* left knee	0 (63, 26%), 1 (20, 8%), 2 (105, 43%), 3 (58, 23%), 4 (1, 0%)	NA	NA
Kellgren- Lawrence* right knee	0 (43, 7%), 1 (18, 7%), 2 (127, 52%), 3 (57, 23%), 4 (1, 0%)	NA	NA
KL* >2 both knees	No (145, 59%), Yes (101, 41%)	NA	NA
KL* >2 either knee	No (0, 0%), Yes (248, 100%)	NA	NA

* 0: normal, 1: doubtful, 2: definite osteophyte (bony protuberance), 3: joint space narrowing (with or without osteophyte), 4: joint deformity

Sample Selection and Pooling Strategies – Replication Sample 2

[0215] A third case control sample (replication sample #2) was created by using individuals with symptoms of OA. These individuals were recruited and examined by rheumatologists. Affected joints were x-rayed and a diagnosis of definite or probable OA was made according to American College of Rheumatology criteria by a single rheumatologist. Controls were recruited from volunteers without any symptoms from the musculoskeletal system based on a normal joint exam performed by a rheumatologist. Only cases with a diagnosis of definite OA were included in the study. Only individuals of Caucasian origin were included. The cases consisted of 228 individuals with definite knee OA, 106 individuals with definite hip OA, and 74 individuals with hip OA.

TABLE 7

Phenotype	Case	Control
Age at Visit	62.7	52.5
Sex (Female/Male)	227/119	174/101
Knee OA Xray: No	35% (120)	80% (16)
Unknown	1% (4)	0% (0)
Yes	64% (221)	20% (4)
Hip OA Xray: No	63% (215)	80% (16)
Unknown	2% (7)	0% (0)
Yes	35% (121)	20% (4)

Assay for Verifying, Allelotyping, and Genotyping SNPs

[0216] Genotyping of the replication cohorts described in Tables 6 and 7 was performed using the same methods used for the original genotyping, as described herein. A MassARRAY™ system (Sequenom, Inc.) was utilized to perform SNP genotyping in a high-throughput fashion. This genotyping platform was complemented by a homogeneous, single-tube assay method (hME™ or homogeneous MassEXTEND™ (Sequenom, Inc.)) in which two genotyping primers anneal to and amplify a genomic target surrounding a polymorphic site of interest. A third primer (the MassEXTEND™ primer), which is complementary to the amplified target up to but not including the polymorphism, was then enzymatically extended one or a few bases through the polymorphic site and then terminated.

[0217] For each polymorphism, SpectroDESIGNER™ software (Sequenom, Inc.) was used to generate a set of PCR primers and a MassEXTEND™ primer which were used to genotype the

polymorphism. Other primer design software could be used or one of ordinary skill in the art could manually design primers based on his or her knowledge of the relevant factors and considerations in designing such primers. Table 3 shows PCR primers and Table 4 shows extension probes used for analyzing (*e.g.*, genotyping) polymorphisms in the replication cohorts. The initial PCR amplification reaction was performed in a 5 μ l total volume containing 1X PCR buffer with 1.5 mM MgCl₂ (Qiagen), 200 μ M each of dATP, dGTP, dCTP, dTTP (Gibco-BRL), 2.5 ng of genomic DNA, 0.1 units of HotStar DNA polymerase (Qiagen), and 200 nM each of forward and reverse PCR primers specific for the polymorphic region of interest.

[0218] Samples were incubated at 95°C for 15 minutes, followed by 45 cycles of 95°C for 20 seconds, 56°C for 30 seconds, and 72°C for 1 minute, finishing with a 3 minute final extension at 72°C. Following amplification, shrimp alkaline phosphatase (SAP) (0.3 units in a 2 μ l volume) (Amersham Pharmacia) was added to each reaction (total reaction volume was 7 μ l) to remove any residual dNTPs that were not consumed in the PCR step. Samples were incubated for 20 minutes at 37°C, followed by 5 minutes at 85°C to denature the SAP.

[0219] Once the SAP reaction was complete, a primer extension reaction was initiated by adding a polymorphism-specific MassEXTEND™ primer cocktail to each sample. Each MassEXTEND™ cocktail included a specific combination of dideoxynucleotides (ddNTPs) and deoxynucleotides (dNTPs) used to distinguish polymorphic alleles from one another. Methods for verifying, allelotyping and genotyping SNPs are disclosed, for example, in U.S. Pat. No. 6,258,538, the content of which is hereby incorporated by reference. In Table 7, ddNTPs are shown and the fourth nucleotide not shown is the dNTP.

[0220] The MassEXTEND™ reaction was performed in a total volume of 9 μ l, with the addition of 1X ThermoSequenase buffer, 0.576 units of ThermoSequenase (Amersham Pharmacia), 600 nM MassEXTEND™ primer, 2 mM of ddATP and/or ddCTP and/or ddGTP and/or ddTTP, and 2 mM of dATP or dCTP or dGTP or dTTP. The deoxy nucleotide (dNTP) used in the assay normally was complementary to the nucleotide at the polymorphic site in the amplicon. Samples were incubated at 94°C for 2 minutes, followed by 55 cycles of 5 seconds at 94°C, 5 seconds at 52°C, and 5 seconds at 72°C.

[0221] Following incubation, samples were desalted by adding 16 μ l of water (total reaction volume was 25 μ l), 3 mg of SpectroCLEAN™ sample cleaning beads (Sequenom, Inc.) and allowed to incubate for 3 minutes with rotation. Samples were then robotically dispensed using a piezoelectric dispensing device (SpectroJET™ (Sequenom, Inc.)) onto either 96-spot or 384-spot silicon chips containing a matrix that crystallized each sample (SpectroCHIP™ (Sequenom, Inc.)). Subsequently, MALDI-TOF mass

spectrometry (Biflex and Autoflex MALDI-TOF mass spectrometers (Bruker Daltonics) can be used) and SpectroTYPER RT™ software (Sequenom, Inc.) were used to analyze and interpret the SNP genotype for each sample.

Genetic Analysis

[0222] Genotyping results for replication cohorts #1 and #2 are provided in Tables 8 and 9, respectively.

TABLE 8

rsID	Replication #1 (Mixed Male/Female cases and Female controls)				Meta-analysis Disc. + Rep #1
	AF OA Con	AF OA Cas	Delta	P-value	P-value
rs1484086	0.23	0.20	0.03	0.287	0.0077

TABLE 9

rsID	Replication #2 (Newfoundland) (Male/Female cases and controls)				Meta-analysis Disc. + Rep #2
	AF OA Con	AF OA Cas	Delta	P-value	Not Done
rs1484086	0.23	0.21	0.026	0.280	

[0223] To combine the evidence for association from multiple sample collections, a meta-analysis procedure was employed. The allele frequencies were compared between cases and controls within the discovery sample, as well as within the replication cohort #1 using the DerSimian-Laird approach (DerSimonian, R. and N. Laird. 1986. Meta-analysis in clinical trials. Control Clin Trials 7: 177-188.)

[0224] The absence of a statistically significant association in one or more of the replication cohorts should not be interpreted as minimizing the value of the original finding. There are many reasons why a biologically derived association identified in a sample from one population would not replicate in a sample from another population. The most important reason is differences in population history. Due to bottlenecks and founder effects, there may be common disease predisposing alleles present in one population that are relatively rare in another, leading to a lack of association in the candidate region. Also, because common diseases such as arthritis-related disorders are the result of susceptibilities in many genes and many environmental risk factors, differences in population-specific genetic and environmental backgrounds could mask the effects of a biologically relevant allele. For these and other reasons, statistically strong results in the original, discovery sample that did not replicate in one or more of the replication samples may be further evaluated in additional replication cohorts and experimental systems.

Example 4

TM7SF3 Region Proximal SNPs

[0225] It has been discovered that SNP rs1484086 in *TM7SF3* is associated with occurrence of osteoarthritis in subjects. *TM7SF3* is an orphan receptor and is a member of the superfamily of 7-transmembrane domain proteins, one of the largest superfamilies of cell surface proteins. Members of this family include receptors for a variety of ligands, such as peptides, hormones, and ions, and for external sensory stimuli, such as odorants and light. Many 7-transmembrane molecules are able to recruit small G proteins, suggesting that they can transduce external signals to the cytoplasm.

[0226] Thirty-seven additional allelic variants proximal to rs1484086 were identified and subsequently allelotyped in osteoarthritis case and control sample sets as described in Examples 1 and 2. The polymorphic variants are set forth in Table 10. The chromosome positions provided in column four of Table 10 are based on Genome "Build 34" of NCBI's GenBank.

TABLE 10

dbSNP rs#	Chromosome	Position in SEQ ID NO: 1	Chromosome Position	Allele Variants
rs1058701	12	230	27004780	a/c
rs11613	12	231	27004781	c/g
rs900743	12	5330	27009880	g/t
rs2129091	12	6334	27010884	a/c
rs15556	12	11372	27015922	c/t
rs1053724	12	11456	27016006	g/t
rs9699	12	11501	27016051	a/g
rs4856	12	13393	27017943	c/t
rs3782314	12	16666	27021216	a/g
rs2087736	12	17596	27022146	c/t
rs2068372	12	19710	27024260	a/g
rs2068371	12	19800	27024350	a/g
rs1184529	12	20297	27024847	a/g
rs2101256	12	20967	27025517	g/t
rs1552257	12	32514	27037064	c/t
rs1565585	12	33159	27037709	a/g
rs4080800	12	37600	27042150	a/g
rs1388658	12	41259	27045809	a/g
rs3782316	12	41329	27045879	c/t
rs1484086	12	50060	27054610	c/t
rs1484087	12	53292	27057842	c/t
rs2291549	12	53393	27057943	a/g
rs1565584	12	56417	27060967	c/t
rs3071166	12	56435	27060985	-/tt
rs1907652	12	58847	27063397	a/t
rs3759119	12	59595	27064145	c/t
rs3759120	12	59661	27064211	g/t

dbSNP rs#	Chromosome	Position in SEQ ID NO: 1	Chromosome Position	Allele Variants
rs3782317	12	60355	27064905	c/t
rs3825166	12	60407	27064957	a/c
rs1872191	12	62357	27066907	a/g
rs8628	12	68230	27072780	g/t
rs1872193	12	68516	27073066	a/g
rs1042833	12	69055	27073605	c/t
rs2476	12	72603	27077153	c/g
rs1976206	12	73928	27078478	a/g
rs1388659	12	85897	27090447	c/t
rs1184528	12	91554	27096104	a/g

Assay for Verifying and Allelotyping SNPs

[0227] The methods used to verify and allelotype the 37 proximal SNPs of Table 10 are the same methods described in Examples 1 and 2 herein. The primers and probes used in these assays are provided in Table 11 and Table 12, respectively.

TABLE 11

dbSNP rs#	Forward PCR primer	Reverse PCR primer
rs1058701	ACGTTGGATGCTTCTCTCCAGTCCATGTTG	ACGTTGGATGGTAACAAGTCCGAAGGATTC
rs11613	ACGTTGGATGTCCAAGTGCCTTCTCTCCAG	ACGTTGGATGTCCCTTGATGCATCTCGACAC
rs900743	ACGTTGGATGAACAGATCCGGAACCTTTTT	ACGTTGGATGGGCTCATAGAACCCCTTTTT
rs2129091	ACGTTGGATGTTAAATTTGGTATGGTCTCC	ACGTTGGATGGTTTTCCTACTAAGAAGAGAC
rs15556	ACGTTGGATGTGTAGGGACAAAGTTATATGG AA	ACGTTGGATGTGTCTTTACAGCTTCACTGGC A
rs1053724	ACGTTGGATGCCATATAACTTTGTCCCTAC	ACGTTGGATGAAAATACAGTGCAGGTGACC
rs9699	ACGTTGGATGTTTACAGTGTACTGAAGTTAAT T	ACGTTGGATGCAACATAGAAAATATAAACAA TTT
rs4856	ACGTTGGATGTTGTAACCAAAAGGTAATTTT TCA	ACGTTGGATGGATGACCAAAAGGTTGGTGG
rs3782314	ACGTTGGATGAATGCTGCACAACCTTAGGCC	ACGTTGGATGAGCAGTCAGCCTATTCTTGC
rs2087736	ACGTTGGATGCAGTGAGACTCTGTCTCTAC	ACGTTGGATGTACATCAGCCTCCCAAGTAG
rs2068372	ACGTTGGATGCTAGAGAGGGTTGCTCTATG	ACGTTGGATGCCAGCTTCAGCTTTCTGTAG
rs2068371	ACGTTGGATGATTGACAAGCCAAAGCCTGTC	ACGTTGGATGATAGAGCAACCCCTCTAGG
rs1184529	ACGTTGGATGTTCTTTCTGGTGGTTTGC	ACGTTGGATGGGTGTCAGTCTTGTAAATCAG
rs2101256	ACGTTGGATGCTGGGTTTCTTTGGTGTGTTG	ACGTTGGATGCTGTGACTGTATCTCAATTTC
rs1552257	ACGTTGGATGCAATACCTTAGCTCACTAC	ACGTTGGATGCATTAGGGAAAATCTGGATC
rs1565585	ACGTTGGATGGCTAGAATATACCACAGGTG	ACGTTGGATGTCACTGAAAATGGATCAGCC
rs4080800	ACGTTGGATGGGGTTTCAAGCAATTTCTCCTG	ACGTTGGATGAAAATTAGCCAGGTGTGGTG
rs1388658	ACGTTGGATGGCAAGAGAGAGGATGGGTG	ACGTTGGATGCCAGGCTTAAAGAACCACTG
rs3782316	ACGTTGGATGGTCTTGCTGTAATGTCAATTAG	ACGTTGGATGCTTCTCTTTTAACTAGATC
rs1484086	ACGTTGGATGTGTCACTCTCCGGAAGTCTC	ACGTTGGATGCATGTACAGGGCAATTCACAG
rs1484087	ACGTTGGATGGCATGGGAATTTTACACCCCT	ACGTTGGATGTCCCTCTCAAACTCTGTAG
rs2291549	ACGTTGGATGACTAAAGCGCCTCTTTGCTG	ACGTTGGATGCATTAAAGCCGCTAGTCCAC
rs1565584	ACGTTGGATGAAAAGGCAAGTTATGATGC	ACGTTGGATGAGTATCCATTCCTCTGAGTC

dbSNP rs#	Forward PCR primer	Reverse PCR primer
rs3071166	ACGTTGGATGGAAAGGCAAGTTATGATGC	ACGTTGGATGAGTATCCATTCTCTGAGTC
rs1907652	ACGTTGGATGCCATATTTCCAGTACCTAGC	ACGTTGGATGGGTTGATAAATATGTGCCAG
rs3759119	ACGTTGGATGCTAAACTGTTCTTAGTGCCG	ACGTTGGATGGGGAATTTCAAAGAGTGTGG
rs3759120	ACGTTGGATGCGGCACCTAAGAACAGTTAGA	ACGTTGGATGGGTTTATGACTGTAGCAAC
rs3782317	ACGTTGGATGGGAAGCTCTTGAAGCTGTAG	ACGTTGGATGGCCATTGAGAAATCTGAGC
rs3825166	ACGTTGGATGACCAGACTCTCAACTTAGCC	ACGTTGGATGTGGCAGGTTGGTTATCTTG
rs1872191	ACGTTGGATGTTTTATGCCGAAGCCCTGTCT	ACGTTGGATGATAAGCAAGGGAGGAAAGGG
rs8628	ACGTTGGATGCAAGCCTTACCAACAATTACA GAA	ACGTTGGATGGCTTATCAAGAGTGAAAATAGA AGA
rs1872193	ACGTTGGATGAATGCCACCTCTAAGAGGCA	ACGTTGGATGCTCAAGCCAAAGGAGAAGAC
rs1042833	ACGTTGGATGTCATCCCAACCTTGGATCTC	ACGTTGGATGATACAGCCCCAGGAAATAGC
rs2476	ACGTTGGATGAGGAGCAGGTGACGTTAATG	ACGTTGGATGATTGACTAGCCACCAGGAAG
rs1976206	ACGTTGGATGAGCTGGGATTGGATTACAGG	ACGTTGGATGATGGAGAAACCCCGTCTCTA
rs1388659	ACGTTGGATGACCTCAGCCTTCCACATTTG	ACGTTGGATGAGCAGGAGGAAACCTTTTGGG
rs1184528	ACGTTGGATGTGAGAATCGCTTGAACCCAG	ACGTTGGATGTTGAATCAGAGTCTCCCTC

TABLE 12

dbSNP rs#	Extend Primer	Term Mix
rs1058701	CCAGTCCATGTTGAGGTGC	ACT
rs11613	CCAGTCCATGTTGAGGTG	ACT
rs900743	ATCCGGAACCTTTTTTTGAAATT	ACT
rs2129091	ATTTAATTTGAAAAGTGCTTACCC	ACT
rs15556	GTTTGTTATTTCTATAC	ACG
rs1053724	CATCAAGTTTCAGTGTTA	CGT
rs9699	TACAGAATAAGCAGTAAA	ACG
rs4856	CAAGTATTTGAAAATAAG	ACT
rs3782314	ACAAATATATGAGAACTCCTCTTT	ACT
rs2087736	ACTCTGTCTCTACTAAAAATAAAA	ACT
rs2068372	TGGCTTCATGGCACCAGCTG	ACT
rs2068371	AGCCAAGCCTGTCACTGGCCCT	ACT
rs1184529	TCCTGGTGGTTTGCCACTTA	ACT
rs2101256	GTTTTTTTGGTGTGTTGATATGTA	ACT
rs1552257	AGCTCACCTACGAAAAATGAATAA	ACG
rs1565585	TGTCAGACTGCACACTACAT	ACT
rs4080800	ATTCTCTGTCTCAGCCCTCC	ACG
rs1388658	GAGAGGATGGGTGAAATAAGG	ACT
rs3782316	TGTAATGTCATTAGGAAGAAACA	ACG
rs1484086	CTCTTCGGAAGTCTCTTCTCA	ACT
rs1484087	TTTACACCCCAAACTAGAG	ACT
rs2291549	CCTCTTGTCTGCCAGTGG	ACT
rs1565584	TGATGCAATAAGTATATATAGTAC	ACT
rs3071166	TAGTACGTGGCTTTTTTTTTTTTT	CGT
rs1907652	TCCAGTACCTAGCACCTAAC	CGT
rs3759119	TTAGTGCCGGATGAATAACT	ACT

dbSNP rs#	Extend Primer	Term Mix
rs3759120	AAATTGCAACTGTGAGTATTAAG	ACT
rs3782317	AGAATAACCCACCTGCCAAAT	ACT
rs3825166	ACTTAGCCTACATTGAAAAGGG	ACT
rs1872191	CTCAGTCCGCTCCCCACTT	ACG
rs8628	TTAAACACTATGACACAT	ACG
rs1872193	AGAGGCAGGACACTAGCC	ACG
rs1042833	CCCTTGGATCTCTTTGAG	ACG
rs2476	GTGACGTTAATGGGACAGCT	ACT
rs1976206	TACAGGCGCCACCACCA	ACG
rs1388659	CCACATTTGGTAAGTTTGACAT	ACT
rs1184528	AGAGGTTGCAGAGACCAAGATC	ACT

Genetic Analysis

[0228] Allelotyping results from the discovery cohort are shown for cases and controls in Table 13. The allele frequency for the A2 allele is noted in the fifth and sixth columns for osteoarthritis case pools and control pools, respectively, where "AF" is allele frequency. The allele frequency for the A1 allele can be easily calculated by subtracting the A2 allele frequency from 1 (A1 AF = 1-A2 AF). For example, the SNP rs900743 has the following case and control allele frequencies: case A1 (G) = 0.92; case A2 (T) = 0.08; control A1 (G) = 0.90; and control A2 (T) = 0.10, where the nucleotide is provided in parenthesis. Some SNPs are labeled "untyped" because of failed assays.

TABLE 13

dbSNP rs#	Position in SEQ ID NO: 1	Chromosome Position	A1/A2 Allele	F A2 Case AF	F A2 Control AF	F p- Value
rs1058701	230	27004780	A/C			
rs11613	231	27004781	C/G			
rs900743	5330	27009880	G/T	0.08	0.10	0.381
rs2129091	6334	27010884	A/C	0.19	0.20	0.646
rs15556	11372	27015922	C/T	0.74	0.74	0.923
rs1053724	11456	27016006	G/T	0.14	0.15	0.592
rs9699	11501	27016051	A/G	0.67	0.68	0.595
rs4856	13393	27017943	C/T	0.10	0.09	0.474
rs3782314	16666	27021216	A/G	0.20	0.22	0.278
rs2087736	17596	27022146	C/T			
rs2068372	19710	27024260	A/G	0.11	0.10	0.634
rs2068371	19800	27024350	A/G			
rs1184529	20297	27024847	A/G			
rs2101256	20967	27025517	G/T			
rs1552257	32514	27037064	C/T	0.67	0.67	0.969
rs1565585	33159	27037709	A/G	0.69	0.65	0.064
rs4080800	37600	27042150	A/G			
rs1388658	41259	27045809	A/G	0.35	0.38	0.362
rs3782316	41329	27045879	C/T	0.47	0.46	0.740
rs1484086	50060	27054610	C/T			

dbSNP rs#	Position in SEQ ID NO: 1	Chromosome Position	A1/A2 Allele	F A2 Case AF	F A2 Control AF	F p- Value
rs1484087	53292	27057842	C/T	0.35	0.38	0.127
rs2291549	53393	27057943	A/G			
rs1565584	56417	27060967	C/T	0.66	0.69	0.327
rs3071166	56435	27060985	-T/T	0.52	0.47	0.042
rs1907652	58847	27063397	A/T	0.64	0.64	0.984
rs3759119	59595	27064145	C/T	0.08	0.05	0.042
rs3759120	59661	27064211	G/T	0.15	0.14	0.861
rs3782317	60355	27064905	C/T	0.08	0.07	0.780
rs3825166	60407	27064957	A/C	0.83	0.84	0.660
rs1872191	62357	27066907	A/G	0.04	0.05	0.555
rs8628	68230	27072780	G/T	0.59	0.61	0.411
rs1872193	68516	27073066	A/G	0.92	0.91	0.669
rs1042833	69055	27073605	C/T	0.85	0.86	0.605
rs2476	72603	27077153	C/G	0.81	0.82	0.392
rs1976206	73928	27078478	A/G	0.16	0.15	0.781
rs1388659	85897	27090447	C/T	0.49	0.50	0.692
rs1184528	91554	27096104	A/G	0.20	0.17	0.228

[0229] The *TM7SF3* proximal SNPs were also allelotyped in the replication cohorts using the methods described herein and the primers provided in Tables 11 and 12. The replication allelotyping results for replication cohort #1 and replication cohort #2 are provided in Tables 14 and 15, respectively.

TABLE 14

dbSNP rs#	Position in SEQ ID NO: 1	Chromosome Position	A1/A2 Allele	F A2 Case AF	F A2 Control AF	F p- Value
rs1058701	230	27004780	A/C			
rs11613	231	27004781	C/G			
rs900743	5330	27009880	G/T	0.07	0.11	0.258
rs2129091	6334	27010884	A/C	0.18	0.20	0.450
rs15556	11372	27015922	C/T	0.73	0.74	0.858
rs1053724	11456	27016006	G/T	0.15	0.15	0.969
rs9699	11501	27016051	A/G	0.65	0.68	0.290
rs4856	13393	27017943	C/T	0.11	0.10	0.735
rs3782314	16666	27021216	A/G	0.18	0.22	0.171
rs2087736	17596	27022146	C/T			
rs2068372	19710	27024260	A/G	0.11	0.11	0.867
rs2068371	19800	27024350	A/G			
rs1184529	20297	27024847	A/G			
rs2101256	20967	27025517	G/T			
rs1552257	32514	27037064	C/T	0.70	0.66	0.239
rs1565585	33159	27037709	A/G	0.67	0.65	0.599
rs4080800	37600	27042150	A/G			
rs1388658	41259	27045809	A/G	0.36	NA	NA
rs3782316	41329	27045879	C/T	0.45	0.49	0.230
rs1484086	50060	27054610	C/T			
rs1484087	53292	27057842	C/T	0.32	0.40	0.023
rs2291549	53393	27057943	A/G			
rs1565584	56417	27060967	C/T	0.61	0.65	0.297
rs3071166	56435	27060985	-T/T	0.55	0.45	0.003
rs1907652	58847	27063397	A/T	0.62	NA	0.638
rs3759119	59595	27064145	C/T	0.09	0.04	0.025
rs3759120	59661	27064211	G/T	0.16	0.15	0.651

dbSNP rs#	Position in SEQ ID NO: 1	Chromosome Position	A1/A2 Allele	F A2 Case AF	F A2 Control AF	F p- Value
rs3782317	60355	27064905	C/T	0.07	0.07	0.986
rs3825166	60407	27064957	A/C	0.84	0.84	0.945
rs1872191	62357	27066907	A/G	0.04	0.06	0.562
rs8628	68230	27072780	G/T	0.58	0.64	0.060
rs1872193	68516	27073066	A/G	0.92	0.91	0.957
rs1042833	69055	27073605	C/T	0.85	0.86	0.746
rs2476	72603	27077153	C/G	0.81	0.84	0.397
rs1976206	73928	27078478	A/G	0.17	0.16	0.788
rs1388659	85897	27090447	C/T	0.47	0.50	0.283
rs1184528	91554	27096104	A/G	0.21	0.16	0.170

TABLE 15

dbSNP rs#	Position in SEQ ID NO: 1	Chromosome Position	A1/A2 Allele	F A2 Case AF	F A2 Control AF	F p- Value
rs1058701	230	27004780	A/C			
rs11613	231	27004781	C/G			
rs900743	5330	27009880	G/T	0.10	0.09	0.763
rs2129091	6334	27010884	A/C	0.21	0.20	0.896
rs15556	11372	27015922	C/T	0.75	0.75	0.974
rs1053724	11456	27016006	G/T	0.12	0.15	0.373
rs9699	11501	27016051	A/G	0.69	0.67	0.615
rs4856	13393	27017943	C/T	0.10	0.08	0.413
rs3782314	16666	27021216	A/G	0.22	0.22	0.944
rs2087736	17596	27022146	C/T			
rs2068372	19710	27024260	A/G	0.11	0.10	0.561
rs2068371	19800	27024350	A/G			
rs1184529	20297	27024847	A/G			
rs2101256	20967	27025517	G/T			
rs1552257	32514	27037064	C/T	0.63	0.68	0.247
rs1565585	33159	27037709	A/G	0.73	0.66	0.029
rs4080800	37600	27042150	A/G			
rs1388658	41259	27045809	A/G	0.33	0.38	0.217
rs3782316	41329	27045879	C/T	0.51	0.42	0.036
rs1484086	50060	27054610	C/T			
rs1484087	53292	27057842	C/T	0.38	0.36	0.575
rs2291549	53393	27057943	A/G			
rs1565584	56417	27060967	C/T	0.73	0.74	0.635
rs3071166	56435	27060985	T/T	0.48	0.49	0.727
rs1907652	58847	27063397	A/T	0.66	-0.02	
rs3759119	59595	27064145	C/T	0.06	0.05	0.669
rs3759120	59661	27064211	G/T	0.13	0.14	0.881
rs3782317	60355	27064905	C/T	0.08	0.07	0.655
rs3825166	60407	27064957	A/C	0.82	0.84	0.546
rs1872191	62357	27066907	A/G	0.04	0.04	0.894
rs8628	68230	27072780	G/T	0.61	0.56	0.234
rs1872193	68516	27073066	A/G	0.93	0.91	0.528
rs1042833	69055	27073605	C/T	0.85	0.86	0.682
rs2476	72603	27077153	C/G	0.80	0.81	0.829
rs1976206	73928	27078478	A/G	0.14	0.14	0.834
rs1388659	85897	27090447	C/T	0.51	0.48	0.499
rs1184528	91554	27096104	A/G	0.18	0.18	0.993

[0230] Allelotyping results were considered particularly significant with a calculated p-value of less than or equal to 0.05 for allelotyping results. These values are indicated in bold. The allelotyping p-values

were plotted in Figure 1 for the discovery cohort. The position of each SNP on the chromosome is presented on the x-axis. The y-axis provides the negative logarithm (base 10) of the p-value comparing the estimated allele in the case group to that of the control group. The minor allele frequency of the control group for each SNP designated by an X or other symbol on the graphs in Figure 1 can be determined by consulting Table 13. For example, the left-most X on the left graph is at position 27004780. By proceeding down the Table from top to bottom and across the graphs from left to right the allele frequency associated with each symbol shown can be determined.

[0231] To aid the interpretation, multiple lines have been added to the graph. The broken horizontal lines are drawn at two common significance levels, 0.05 and 0.01. The vertical broken lines are drawn every 20kb to assist in the interpretation of distances between SNPs. Two other lines are drawn to expose linear trends in the association of SNPs to the disease. The light gray line (or generally bottom-most curve) is a nonlinear smoother through the data points on the graph using a local polynomial regression method (W.S. Cleveland, E. Grosse and W.M. Shyu (1992) Local regression models. Chapter 8 of Statistical Models in S eds J.M. Chambers and T.J. Hastie, Wadsworth & Brooks/Cole.). The black line provides a local test for excess statistical significance to identify regions of association. This was created by use of a 10kb sliding window with 1kb step sizes. Within each window, a chi-square goodness of fit test was applied to compare the proportion of SNPs that were significant at a test wise level of 0.01, to the proportion that would be expected by chance alone (0.05 for the methods used here). Resulting p-values that were less than 10^{-8} were truncated at that value.

[0232] The exons and introns of the genes in the covered region are plotted below each graph at the appropriate chromosomal positions. The gene boundary is indicated by the broken horizontal line. The exon positions are shown as thick, unbroken bars. An arrow is placed at the 3' end of each gene to show the direction of transcription.

Example 5

In Vitro Production of Target Polypeptides

[0233] cDNA is cloned into a pIVEX 2.3-MCS vector (Roche Biochem) using a directional cloning method. A cDNA insert is prepared using PCR with forward and reverse primers having 5' restriction site tags (in frame) and 5-6 additional nucleotides in addition to 3' gene-specific portions, the latter of which is typically about twenty to about twenty-five base pairs in length. A Sal I restriction site is introduced by the forward primer and a Sma I restriction site is introduced by the reverse primer. The ends of PCR products are cut with the corresponding restriction enzymes (*i.e.*, Sal I and Sma I) and the products are gel-purified. The pIVEX 2.3-MCS vector is linearized using the same restriction enzymes, and the fragment with the correct sized fragment is isolated by gel-purification. Purified PCR product is ligated into the linearized pIVEX 2.3-MCS vector and *E. coli* cells transformed for plasmid

amplification. The newly constructed expression vector is verified by restriction mapping and used for protein production.

[0234] *E. coli* lysate is reconstituted with 0.25 ml of Reconstitution Buffer, the Reaction Mix is reconstituted with 0.8 ml of Reconstitution Buffer; the Feeding Mix is reconstituted with 10.5 ml of Reconstitution Buffer; and the Energy Mix is reconstituted with 0.6 ml of Reconstitution Buffer. 0.5 ml of the Energy Mix was added to the Feeding Mix to obtain the Feeding Solution. 0.75 ml of Reaction Mix, 50 μ l of Energy Mix, and 10 μ g of the template DNA is added to the *E. coli* lysate.

[0235] Using the reaction device (Roche Biochem), 1 ml of the Reaction Solution is loaded into the reaction compartment. The reaction device is turned upside-down and 10 ml of the Feeding Solution is loaded into the feeding compartment. All lids are closed and the reaction device is loaded into the RTS500 instrument. The instrument is run at 30°C for 24 hours with a stir bar speed of 150 rpm. The pIVEX 2.3 MCS vector includes a nucleotide sequence that encodes six consecutive histidine amino acids on the C-terminal end of the target polypeptide for the purpose of protein purification. Target polypeptide is purified by contacting the contents of reaction device with resin modified with Ni^{2+} ions. Target polypeptide is eluted from the resin with a solution containing free Ni^{2+} ions.

Example 6

Cellular Production of Target Polypeptides

[0236] Nucleic acids are cloned into DNA plasmids having phage recombination sites and target polypeptides are expressed therefrom in a variety of host cells. Alpha phage genomic DNA contains short sequences known as attP sites, and *E. coli* genomic DNA contains unique, short sequences known as attB sites. These regions share homology, allowing for integration of phage DNA into *E. coli* via directional, site-specific recombination using the phage protein Int and the *E. coli* protein IHF. Integration produces two new att sites, L and R, which flank the inserted prophage DNA. Phage excision from *E. coli* genomic DNA can also be accomplished using these two proteins with the addition of a second phage protein, Xis. DNA vectors have been produced where the integration/excision process is modified to allow for the directional integration or excision of a target DNA fragment into a backbone vector in a rapid *in vitro* reaction (Gateway™ Technology (Invitrogen, Inc.)).

[0237] A first step is to transfer the nucleic acid insert into a shuttle vector that contains attL sites surrounding the negative selection gene, *ccdB* (e.g. pENTER vector, Invitrogen, Inc.). This transfer process is accomplished by digesting the nucleic acid from a DNA vector used for sequencing, and to ligate it into the multicloning site of the shuttle vector, which will place it between the two attL sites while removing the negative selection gene *ccdB*. A second method is to amplify the nucleic acid by the polymerase chain reaction (PCR) with primers containing attB sites. The amplified fragment then is

integrated into the shuttle vector using Int and IHF. A third method is to utilize a topoisomerase-mediated process, in which the nucleic acid is amplified via PCR using gene-specific primers with the 5' upstream primer containing an additional CACC sequence (e.g., TOPO® expression kit (Invitrogen, Inc.)). In conjunction with Topoisomerase I, the PCR amplified fragment can be cloned into the shuttle vector via the attL sites in the correct orientation.

[0238] Once the nucleic acid is transferred into the shuttle vector, it can be cloned into an expression vector having attR sites. Several vectors containing attR sites for expression of target polypeptide as a native polypeptide, N-fusion polypeptide, and C-fusion polypeptides are commercially available (e.g., pDEST (Invitrogen, Inc.)), and any vector can be converted into an expression vector for receiving a nucleic acid from the shuttle vector by introducing an insert having an attR site flanked by an antibiotic resistant gene for selection using the standard methods described above. Transfer of the nucleic acid from the shuttle vector is accomplished by directional recombination using Int, IHF, and Xis (LR clonease). Then the desired sequence can be transferred to an expression vector by carrying out a one hour incubation at room temperature with Int, IHF, and Xis, a ten minute incubation at 37°C with proteinase K, transforming bacteria and allowing expression for one hour, and then plating on selective media. Generally, 90% cloning efficiency is achieved by this method. Examples of expression vectors are pDEST 14 bacterial expression vector with att7 promoter, pDEST 15 bacterial expression vector with a T7 promoter and a N-terminal GST tag, pDEST 17 bacterial vector with a T7 promoter and a N-terminal polyhistidine affinity tag, and pDEST 12.2 mammalian expression vector with a CMV promoter and neo resistance gene. These expression vectors or others like them are transformed or transfected into cells for expression of the target polypeptide or polypeptide variants. These expression vectors are often transfected, for example, into murine-transformed adipocyte cell line 3T3-L1, (ATCC), human embryonic kidney cell line 293, and rat cardiomyocyte cell line H9C2.

Nucleotide and Amino Acid Sequence Embodiments

[0239] Table A includes information pertaining to the incident polymorphic variant associated with osteoarthritis identified herein. Public information pertaining to the polymorphism and the genomic sequence that includes the polymorphism are indicated. The genomic sequences identified in Table A may be accessed at the http address www.ncbi.nih.gov/entrez/query.fcgi, for example, by using the publicly available SNP reference number (e.g., rs1484086). The chromosome position refers to the position of the SNP within NCBI's Genome Build 34, which may be accessed at the following http address: www.ncbi.nlm.nih.gov/mapview/map_search.cgi?chr=hum_chr.inf&query=. The "Contig Position" provided in Table A corresponds to a nucleotide position set forth in the contig sequence (see "Contig Accession No."), and designates the polymorphic site corresponding to the SNP reference

number. The sequence containing the polymorphisms also may be referenced by the "Nucleotide Accession No." set forth in Table A. The "Sequence Identification" corresponds to cDNA sequence that encodes associated target polypeptides (e.g., TM7SF3) of the invention. The position of the SNP within the cDNA sequence is provided in the "Sequence Position" column of Table A. If the SNP falls within an exon, the corresponding amino acid position (and amino acid change, if applicable) is provided as well. Also, the allelic variation at the polymorphic site and the allelic variant identified as associated with osteoarthritis is specified in Table A. All nucleotide and polypeptide sequences referenced and accessed by the parameters set forth in Table A are incorporated herein by reference.

Table A

RS_ID	Chromosome	Chrom Position	Contig Accession No. [1]	Contig Position	Nucleotide Accession No. [2]	Sequence Position	Amino Acid Position	Locus	Locus ID	A [3]	Allelic Variability	OA Assoc. Allele
rs1484086	12	27054610	Hs12_9871_34:16	19922317	NM_016551	intron		TM7SF3	51768	R	[T/C]	T

[1] Contig Accession Number which can be found in the NCBI Database;
[http address: www.ncbi.nlm.nih.gov/entrez/query.fcgi](http://www.ncbi.nlm.nih.gov/entrez/query.fcgi)

[2] Sequence Identification or Nucleotide Accession Number which can be found in the NCBI Database;
[http address: www.ncbi.nlm.nih.gov/entrez/query.fcgi](http://www.ncbi.nlm.nih.gov/entrez/query.fcgi)

[3] "A" column is the sequence orientation ("F" is forward, "R" is reverse).

[0240] The following is a genomic nucleotide sequence for an *TM7SF3* region. The following nucleotide representations are used throughout: "A" or "a" is adenosine, adenine, or adenylic acid; "C" or "c" is cytidine, cytosine, or cytidylic acid; "G" or "g" is guanosine, guanine, or guanylic acid; "T" or "t" is thymidine, thymine, or thymidylic acid; and "I" or "i" is inosine, hypoxanthine, or inosinic acid. Exons are indicated in italicized lower case type, introns are depicted in normal text lower case type, and polymorphic sites are depicted in bold upper case type. SNPs are designated by the following convention: "R" represents A or G, "M" represents A or C; "W" represents A or T; "Y" represents C or T; "S" represents C or G; "K" represents G or T; "V" represents A, C or G; "H" represents A, C, or T; "D" represents A, G, or T; "B" represents C, G, or T; and "N" represents A, G, C, or T.

TM7SF3 Genomic Sequence (SEQ ID NO: 1)

```
>12:27004551-27096350
1  acccaaatatc taattttcat gttatttttg aacagatgcc tttttggaaa tcttaaaaaa
61  gcaaacattt caaacagttg gctaacacgc atgaactaac atcatttcag tttttaaat
121  acggatgata tccgtattaa cattataatg actctatact acagattgac atggatcac
181  gtaacaagtc cgaaggattc ttccttgatg catctcgaca catcttgaM Scacctcaac
241  atggactgga gagaaggcac ttggaagcaa atcagaatgt acactaaata aacagttcaac
301  ttttggggtg tgtagggaag ggggggtccat tttaaaagt cttttacatt gaatttcctt
```

361	cccagattag	atcagcaaat	aatgaaatt	tattaaaaag	atgttgtgtc	tttaaatgac
421	atgctgttca	gtgtagaatt	tttaaatctc	tattaaattt	tttgtgtatt	gtatcatgtt
481	cagcctgtga	gatatcttgc	ggttttatag	gttattcagt	ccaaagctgt	atttatttgc
541	ttcttgttagt	tggttgtgtg	aaacagtcac	ttccttctgt	gacatcaaa	atttatttgc
601	ggcattaat	ttaaactgtg	ttcataaggt	caaatgtatt	aattgtttta	ttcatagttg
661	ggatagtttg	tagtcttctg	attttgcasa	tgaatttaa	atgcttttgt	ttatagtttg
721	ggatagtttg	tagtcttctg	ttctccattg	ttcttcaatt	ttttttttac	ttcatagttg
781	caggtctctta	ttgaataaa	gaggaaataa	ccatagacag	gaggttagaa	ttgaagacat
841	ttcaagacag	ttcaaggggt	atgttgtgtg	cacttcttgt	cttccagaca	gttttgaag
901	caagatctct	gtgcttgtgt	cttgttaaaa	aaactgtgat	actctctcga	agattttata
961	attcaataag	tgaaacatta	ccatgacctt	ttataaaagt	gctgcacctat	ttgaagtatt
1021	gggttttttc	tcattgatag	ctttgttaga	cttggttaagt	aaaaaatcag	tagttttttt
1081	cttctatttt	taaaaggcaa	gaaatgctgt	cttcaagctg	gcattggtggc	tcattgctgtg
1141	aattggggag	ctgaggggtg	tgatcacact	gacgtcagga	attggagacc	agcctggcca
1201	acattgtgaa	accctcatct	tactaaaaat	acaaaaaggt	agccagcggt	gatagcggtg
1261	gcctgtgaa	ccagctgtct	gggaggtcaa	agcaagagaa	ttgtttgaa	ctggagatgg
1321	gaggtgtcag	tgagccgagg	tcgtgccatt	gcactccagc	ctggccaaca	agagccagac
1381	ttctgtcaaa	aaaaaagaaa	tgcttgtctt	ctgtgtagtt	ttttaactta	gttcaagtgc
1441	ataggactga	ccaaatcata	aattgaggat	gctgttgtta	ttcaaaagtt	ttactactta
1501	cccttactca	gtttgcagct	gcaatgatct	gtacttaacc	tgcttttaca	tattatgtct
1561	cccttgacaat	gggaggaatt	ttaatggtta	acttgcagaa	tatgacagat	acttgcctta
1621	atatcttctt	tctaccata	atctaatcca	aaactatcca	aggaactctt	ctgaaattgat
1681	tggaattggt	tttactctgt	ctgcccagtg	tggaagccag	tagccaactgt	tggtcttgtg
1741	gcactgtcag	gttacttaatt	gcgtttgagg	gactgaattg	ttaaattttta	aaaaatgtgt
1801	gatttttaatt	aagtttaaca	ttgttttgga	cagccagagat	tagccttttta	gagttacctt
1861	ttgtctgagt	ctagcagttg	ataggtttcc	ccatgaaag	gacctaatgt	ttttgttgtt
1921	aagaggttat	taaatgtgct	ttgccaaga	tactacctgt	tcatttctct	cagcttaata
1981	gttccogtgt	ttcttggagg	tactgttaaa	tcctttttac	aaatataat	atacacaaat
2041	actgaagaga	ctgttttatt	ggggagttat	tcacttttat	atttaagctt	taggcttaatt
2101	atgataaaa	actctcagtt	tgctgtgaat	aatggtcctat	agaagctgtt	catgagaana
2161	atgatatttc	tcagatttct	gttgatttta	aaactttttt	gtctgtctct	ttcatagac
2221	tgtaagcttc	ttcagatcag	gggacatgct	ttgacagctc	ttgtattctc	ttgttgttag
2281	atagtactgt	ccatcatag	gtgtgcaaga	taagaaaaat	agatttggaag	aatttatgct
2341	acttgtctgc	cccaacaaat	ttggatgtct	gtattatact	gtactactat	tttcaatagt
2401	tagcaaaaga	ggccatttaa	tattttaata	ctgaaaagca	gaattgtgct	catttttatt
2461	aaactctacc	atataccata	cactagatag	taggtgtgat	taaaagtcca	gttctgtctc
2521	tcaaaagaga	ttatgacct	gaaaaggtta	cagattggat	tgagataact	ttataggcct
2581	gaacaaatga	gaaaagattt	ctaaatcatt	gaagacttca	gggaggaana	gagattgaaat
2641	gcgtcggaag	tttttatcat	ttccgttact	ttgaacaata	gtgtatgtaa	gtgcaaaaaa
2701	aaagaaaaag	ttgctgatca	tttgtgtgag	aggaatgtaa	aatatgagta	tttttttgtc
2761	tgctaggttt	agttatatgt	tgctgcttat	ttctaatctc	agtcctcaaca	aatttgattt
2821	caaaaagaca	tcgtgtgtca	accacacttc	ttgttaataat	ttgtttttat	caattatagt
2881	gacaatgtaa	gaatatctat	gttaagtgtg	actgcagatt	ttataaaaa	gagaggtacc
2941	aaaaataaaa	aagatttttt	ctgatttcca	actccactcc	cttttttttc	ggagactgca
3001	gcacatgtgt	accagataac	tgaaaatgga	gcagttaata	ggaaaagcat	tgaaattgac
3061	gagcaacagg	gttgcaagga	acaagaacga	atatttcaac	ttgaagttaag	ttttacatgt
3121	caaatgtaga	tgaaaaactgt	coatataaag	atttatgcca	gaaaataat	ccattttttg
3181	aatatatttt	tactcttctt	gggtactggt	aatgttctcc	ttgattgtgt	ctattgtgta
3241	atattatgca	agaccattta	tggtgtgatg	caagaatcta	ggattattag	taacttcaaa
3301	tataccagta	atgtgaaaa	ttgatgtttc	attctatat	gcttatctta	tcagtgtgaa
3361	gtacttcttc	aattgagtca	tttttaaaat	ttgtttttat	ctattagaga	ctcatgtgat
3421	acagggcttt	tttttttaat	ccacaataat	ttaggggcat	ttctattttt	ttcaggtgac
3481	taaaaacctt	tgtcatctgt	cttttatttt	atgacagtga	aagtacaagt	cttcgataaa
3541	ctgaaaatca	tacaatagac	ttttatgcct	taagagctga	attttgaagt	cagggcggtg
3601	aacctgcttt	agactactaa	aatttagaaa	gtgagactga	agataaagt	atagagaaca
3661	aattatgaaa	gggtattgtt	aaatgtctac	aaattttttt	acagtaacta	acaaaagaa
3721	ggccagaaag	caaagtattt	taccagtgtc	agattttggt	acttacttct	cttcttgaaa
3781	aatatcatca	gtcaccctca	ccctgctgct	ctcaaatgca	aattgactca	acttgtccta
3841	catctcatgt	tatcatggaa	agtaataaaa	atactaaaaa	ttactaggag	attctagact
3901	attcactgtc	tttaccagtt	tataataaaa	taaatctaga	tgttattttc	ttcatgctta
3961	atgtgtattg	aaaaaaagcg	aatttctcag	cttctgttta	caattatagc	tgcttaagt
4021	agaagccttt	gcagtggtgt	aagatgttga	acacagttcc	tagctgcata	aggtgtcttt
4081	agattccaagt	tgccaactgc	atatatcatt	gagtagatta	atgccaatta	cagactgtgc
4141	ttgtttgtta	taccttcata	tggttcagtt	ttcctaacat	atgtttgtga	ttgtgtagct

4201	accttgtaga	tagtgactgc	atctttgttt	tgtcttgcc	tttagcaaga	aaacaaaggc
4261	tcggagagaga	tccttcaaat	aactcgagaa	tcatttttga	acctgaagag	gagtagtgcg
4321	tcgagaggtt	ctgtttgtgc	agcattagtg	accaacagtg	atttgagctc	agaggaagc
4381	tcgagaggtt	ctgagctcgt	gagcttctta	catggctcca	atttggtcaa	taagtgaatg
4441	aatgataagg	cgaaatctc	aatctcttat	ttttttctct	gtaaattatg	accagtcacg
4501	ggctatgaga	tagcaacaaa	aaatgcatag	ttaatggctga	tagactttat	tccaaaaatc
4561	aattggaaaa	tagaaactga	gccattgcc	aatggtaaa	aaatgaaa	ttttccagtg
4621	gactactgaa	tataccaaga	ctttttggca	gtactctgtg	ctttctgggt	gatttaattg
4681	gtaaaactga	atattccca	taaatgtttg	agaatgcata	aaattatccc	aattttaaaa
4741	gtaaaatttt	tgcactgtt	tgtgaaagg	aggatttga	cgctaaattt	cactcttatt
4801	tggtcgtaaa	gtgatgcga	agtttatgta	ccatgagact	tctagttagt	gcctactttc
4861	acatttattt	aaatttttgt	gaaagaaaa	acagttttaa	aggctatagg	aatgtgttat
4921	atgtacctta	aacagataat	ttaattttag	caatttatca	taactgatac	ttagctaaac
4981	attgtccata	acttgtttaa	gtgaagaaa	taccatttca	ggaatgaaa	agaaggaata
5041	ctacttttca	agaagaaga	tcctataata	gacatttata	gtaggtttaa	ctactctctt
5101	gaaagaataa	gttttgggtc	aaagtcaata	atgaagacga	gatttgtttt	tctctctctt
5161	ggctgtattt	taaggataat	atgtagtcca	tttagttaac	agaattgaaa	tgtttgatat
5221	ggcaagatag	acttgttaat	ttcaaatgta	attgggaatt	ccctctggctc	agacacccct
5281	tttttttttt	cttttaagat	tcttgagata	caaaaaaaa	aaagtaataM	aattttcaaaa
5341	aaaaagtttc	ggatctgttt	ttaagctcca	tctgtctctc	ataacctgca	agattttttct
5401	taaaaccttt	cagctgaaa	tggtgggttaa	gggtggagtaa	tctgtgtgatt	tgttttctgt
5461	gtcttttttaa	atgtcaata	tataatattg	aattttttta	aaaaaccaca	gatacagaaa
5521	tgtgtctttaa	catcagttga	aacctaaatt	ttcttatgtt	gtgtgtgattg	tattaaaaag
5581	ggatataaag	acagtgctca	acatggttaa	atataattga	ctgaattgat	tgattatttt
5641	attataaatt	agacaaatgg	aaaattatac	tttagtata	taattttgta	aaattacttt
5701	tatatgtgta	ttttatgtat	aatttccata	attggtaaaa	tccaaactca	acttgtagaa
5761	tttttttttc	ttaagtttgg	gggtgaatggg	gtggaatgga	ctgattgaat	agaaaagggtc
5821	taatgtccca	aacattatata	agattttctt	tttccagtca	gaggtcctta	tgtatatttt
5881	ataaataaat	gacagttttt	atttttaaac	tttttttgtt	ttttgggaaa	gtatttcccta
5941	atttataatg	acattcattc	agatacttct	tatccctgct	ataaaggaaa	acttaactta
6001	agctacacaa	ttgagattaa	gtctgaggtca	gttccattgag	gcagctctac	tataaaagct
6061	tacttgataa	ataattattt	ttgtaacaaa	gttgggttaa	cttatttctc	gtctttttgtc
6121	tttgataatga	atttaagggtc	ttcatgttta	aaagacttta	cttttgtatt	tgtagtacaa
6181	tttccatctc	attttttttt	tttttttggt	tgtgtttaa	cagaacctta	agtttattgt
6241	tgaggtatgt	actgcatagg	aacctatttt	attattaaag	atgaattgat	aaaaattgta
6301	tgtgtctccaa	tttaatttga	aaagtgccta	cccKtattct	tatatattgt	ttaatttttaa
6361	ggtttttttg	cttctcttag	tgcaaaacta	cttagcagtg	acctctatct	gtatttcccta
6421	ggaattagca	gcttcttagt	gtggatccctg	cagaacttct	taccatttgt	agtaggttga
6481	atcatgtccc	ctagaaggta	agctcaagtc	ctaacttgat	acacctggga	aggtgacctt
6541	attttgaaat	agcttttaca	gatgtgatta	ggggatctca	ggataataat	accttggttt
6601	ttgggtgggc	tccaaaccca	gtgactagtg	tcgttacaat	aaagagagag	gagatattgaa
6661	acacagata	agcagcatga	ggacagaggg	aatgagttac	gcaaggagctg	ctgaccacaa
6721	ccagaagctg	caagagaggg	atggaatggg	ttcttcacgg	gtctccagaaa	ggaccaagcc
6781	tgtcttgatt	cagacttctg	gcctccagaa	ctgtgaggaa	ataaatttct	gtgttttttc
6841	ttcaccggcg	ctgtggtta	ttgttagagt	aaccttagga	aaactaagca	tctattgttt
6901	ttcatctatc	aactataatt	ttattttata	taagaaagtt	ctttaaacc	aatttataac
6961	ctcaaaattt	ctgctctaat	ttataattct	ttaatatcaa	gggtgactct	aaaaatttat
7021	agaaaataat	tgtctccctg	taatcccgag	actttgggag	gcggaaccag	gtgagtcagt
7081	tgaggtcagg	agttcaagac	cagcctagcc	aaactggtaa	acctgtctc	tactaaaaat
7141	acaaaaggat	agccaggat	gggtgtgggt	gcctgtaatc	ccctgtctct	ggaggccga
7201	ggcaggagaa	tcgcttgaac	ctgggaagca	gaggttgacg	tgagccgaga	ctccaccttt
7261	gcactccagc	ctgggcaata	atagtgaaac	tccattccaa	aaaaaaacaa	aaatttgtct
7321	caatgaaaaa	tggaacaatc	tagatccata	attttgtgtg	tcactacctt	ttttcttgtt
7381	cagagtgaaa	taggagaaca	tttttttagta	aaactctgat	gagatagata	atttctaatt
7441	ggaatttgag	gagacacttt	ttcttttaac	ctctctggct	gtctctcagt	ctccttttgt
7501	tctctctcgt	gaccttttaa	cattagagag	ccactgggct	tagtcttatt	acctattctg
7561	tgactgtctc	ctactcactg	gtgatctcac	ccagtctcat	gactaaatca	caaccttatg
7621	cttgatggct	ctaaaattca	cagttccagc	atagacctct	ccccccggt	tcatgagctg
7681	gtgactgtca	aagagaaatt	ttatgcttaa	gatgtccaaa	gccaactctc	aaacctccca
7741	gaacgctctc	atactcttct	ccatctcgtc	tgatgacaa	tcactctctg	ctactgtctc
7801	ggccaaaaac	cttggtggca	actttgactc	caactttctca	tatcgttgat	ctctatctct
7861	agcaaaatct	gttgttggtc	cttccctcca	aaataaatca	aatctaccca	tttctctctc
7921	ccactgctac	ccctggtcaa	agctacatata	tgtgtattgt	aaagtataca	aattgattct
7981	aaactaaata	tctcacagta	ggggaatagt	taaatgtcaa	tattcaactca	atacagttga

8041	ttacagtcta	tggcttgaag	atgctagcaa	caagaagaaa	aaagtacccta	aatatgcatt
8101	aaaaagctta	acataaatta	cccataaat	tacaaatagt	taaaatagct	atgcatcagg
8161	ggaagtgtcc	tgaagcaaac	agcatcagta	ttctccagc	taattgtact	ttgtgcatta
8221	gccagtcaat	ttctgatgcc	tgtaaatcaa	gcatttltgt	agctaccatt	atggagttgt
8281	aattatctca	tcagttccccc	caactgatgg	atagttttgca	aagttcaata	caaaataagt
8341	cagtaaaacat	ccatatacat	aactttatgt	attgtcaatt	ctagctgca	aatccccaga
8401	agtttgttag	gtcatgtttt	gttatttttt	aaatgaaaaa	ataaatgcaa	actctgtaaa
8461	atctctagcat	actcaaaaaat	ttacaattca	gagtaaacct	ctctctcaaa	ccctctgtct
8521	tcagagagat	attataagtt	ttctgtgaat	cagaaaacct	gttaaccgaaa	aaaagtgaat
8581	gaggccaagt	ttctaaatca	gtagaggttt	ctggggccaa	agttttaaaga	catgccagag
8641	aaaaaacgca	agccacatga	gcactcttga	cttggagcttt	ctaaagaaggg	tttttgagag
8701	tcatatttta	ggggagagag	tgaggagagg	aaagggagag	gggttaggca	gtgaggcaaa
8761	tgattacatt	ctgtgagagg	tctgattaac	gctcagtaaa	tctatgtttc	atataagata
8821	aactgaaaaa	aggagtagaa	gactcagtta	ttcatctgtt	tcagggttag	cagaagaagt
8881	gttgattttg	tctcatcctt	attctgtacc	tggaagata	agcttgttat	cagcatgttc
8941	agtggtgaga	ttaacaaaac	ttggttttaa	gagttaaact	taagttgcag	accaaggttt
9001	acaattgcta	tggtgtgtgt	ttacagctct	ctacaactgt	aaggggaggc	cttgagagct
9061	ctgttaataa	tactcccttt	ctcacacatg	acatacttg	ttttgtgtac	ctctactcaa
9121	ctctccaaagt	tcagctggaa	tgctcaactcc	tcaaggactc	ctctctgtgc	attccccatt
9181	ctagccctat	ccaktatac	attttcaatt	catattacac	actttgttga	actatttttag
9241	tatctctctc	acactccaat	cccagacagc	aaagaccatg	ctgtgttttt	ccattctctt
9301	agtgctcaag	gcaatgctct	gccaggaact	caacaaatct	ttaatgagt	aatgtgagat
9361	ctaaagctaaa	gctctcaatt	catatataata	taatttttat	ttttttgaga	cagttgtctg
9421	ctttgtcgcc	taggtctggag	tgcataatgt	cgactcagtt	taactctgag	gtctgcccc
9481	cggtgtccag	tgattcttcc	tctctagcct	cccaagttag	tgacacagtt	gttccatcat
9541	gccccgtctg	tattttttat	tttttagtag	agacggggtt	tccacagttt	ggccagctgt
9601	gtcttgaact	cttgacttca	agtgatctgc	ccactcttgc	ctctcaaatg	tgctgggata
9661	caggctgtag	ccaccacgcc	actcctgtat	aaccaatttt	taaaagtctt	baactgtgat
9721	ccagacactgc	caactctctc	agtgtaatga	ttacatggta	gtccctgggt	aagtccagag
9781	ttcccactaa	cttaagcccc	ttctgtgttc	aatttcataa	agagtgtcac	tgcgaaatct
9841	aatggcccaa	ttgccaaagag	gcattttcaa	caacagtcag	ggtaattgtt	attctcttgt
9901	cataacagaa	taccactgac	tgaattggctt	aaacaacaga	aaatatttct	ccagttcttg
9961	gagctgtgaag	tccatgatca	aggtgcagca	gggagtgttt	ccctctgttg	cagatgcctc
10021	ctctctactgt	ctcttcacac	ggctcttccc	ctgtgagcac	atgcaccttg	gtttctctgt
10081	tcttaattctc	ccctctctac	aaagacactg	gtcagactgg	ataggagccc	accctaatag
10141	tctcaatttta	atttgaaccc	tctttaaagg	cccctgtctc	aaataaggtc	actattctgag
10201	gtaccgggtg	ttagggcttc	aacataataa	tttgtttggg	gggccaagtt	tagtctatca
10261	gcacaacata	ggtagctgcc	ctcccaacaa	gggctgggat	gcaatgaggg	gctgtgccct
10321	aggcgcaagg	ctgttctcca	ggaaattaga	accaagaact	ctgggtttga	atgaattgta
10381	ctaacaggta	ttaagtacat	cacctaggag	tgaactctgg	accaagtttc	tattgttgag
10441	atatggatatt	tggaaatggt	aaatatattg	taccaatcat	tgaataataa	tggctccatta
10501	cttcagataa	ttaactttgga	aaaacatgct	ttcactgtgt	aaaattataa	attttcaaaa
10561	agaaacgatcg	tagcaataact	tatatagaaa	acatctctac	aattctgttag	cagaagaaaa
10621	tagaataaatac	ctggtttttac	actctgatca	tgtagagtgt	aaagcacatt	cttgatagga
10681	gtcaggctac	ccagtttttg	tctgagcact	gttagttgat	aactctctga	cttcatatac
10741	ttaactctaag	aatttctcca	ggatattaaa	aggtcagtga	acatgacctc	gtgaagctct
10801	tgaagtttcta	atattcttagc	aatttctttg	ttggaatcag	ttttctattt	cactgtctac
10861	agaaaataac	actggcctgt	ggtaacaata	acctagtatg	gtttggcaac	ctgttagcta
10921	acctataaaca	gattaatcag	aactttctga	aaacatttta	aagaacagtt	caacagagtt
10981	tctctgttaac	tactactttt	ctttagtctg	acacacctga	agtgcttaata	accaccaaat
11041	agtccttttaa	cacattgtccc	cagatttcat	ttctctgtat	tatctttgtg	atttttttaa
11101	aatgaagaac	tatatgtttt	gcccatatta	aataacttta	actgatgtga	actctgtcat
11161	gtctttgttgc	tgtgggtacaa	atataattctg	aaaaatttta	tgaataatatt	taattgagct
11221	tttttagcatt	tttttcaaaa	tagggattta	ttataaaaac	tatacaaaact	tttttcaact
11281	cacaaatgat	gaaaaatttt	aaacaattag	tatctctccc	ccaaaataagt	gtgttctctc
11341	tttttagctct	tcactggcag	tgcatccaag	tRgtatagaa	aaatacaaac	attccatata
11401	actttgtccc	tacagtaaat	aatttttttaa	aacttttcat	caagtttcaa	tgttaMtgaa
11461	gttaattatta	gacagtaagg	attacagaat	aagcagtaaa	Yggtcacctg	ctgtgtattt
11521	ttcaaaagcaa	aaattgtttt	atattttcta	tgtgttttta	tctctgcccc	aaactctaaa
11581	gcaaaagttca	aatatttttc	cagttacctg	aaactcaaaa	actctatgtt	actttgaaat
11641	aactttaaaa	agcagcatca	tttaaatagt	ggttaactca	taaacgtgta	tgcacaacca
11701	ttccacacac	ctaataccac	tcaggccctgc	agagtaccac	gtccacaata	attcagtgct
11761	ctgaaagctg	gggggaaagc	ccctgagtag	taagtatgca	accagggtgt	tcgggtgctt
11821	ttctgtgggt	aaacaagggt	gtcaaaaactg	aggggactct	ccaaggtcct	aatgaaaagc

11881	agctcctcta	aacacctgaa	tcctttttga	ctccattctc	atcactttaa	ggatgaagcc
11941	ttctgtgtaa	aaggctttgt	acagaaatta	ttaaaacta	tatttttaac	taagtctaac
12001	agctaactga	gcgaaagaaa	ctatgaagta	caaaattttc	taaggttttc	tatttaaatgg
12061	caatacacag	tttttttttt	tttgaagcgg	agctctcgtc	gttcgcccaag	ggctggagctg
12121	agtggtctga	ttcttggtctc	ctgcaacctc	tgctctctgg	gttcgccaga	tctctctggc
12181	tcagctcccc	gagtagctgg	gattacagcg	acctgcacca	atcccagctg	aattttttgta
12241	tttttagtag	agcaggggtg	tagtagtata	tattctacta	tactatacta	gagaccacta
12301	tgctggccag	gtctggtctcg	aactcctgag	ccacgggtgat	ccaccggcct	cgactcccca
12361	aaatgctggg	attacagagtt	tgagccaacg	cacctggtcg	gcaataatga	ttaaaggtggg
12421	gggtatttcc	taaaaacaaa	agtaatccat	atgtcatcta	atgcagttctg	ataaagctga
12481	actgtgtgac	actattttgaa	aaggcctctga	aaaaaaaaag	gggggcattt	tttggtaat
12541	aaataccctc	tataaaattc	tattactcat	ataaattggc	ctattttcoa	ataataatt
12601	agctacttaa	attaccattt	taagcaataa	aaggcaaaaa	caaagggaac	ctactgaatt
12661	gtcacatttc	agcagaataa	gaaggtaaat	ttcatgtgta	ttttatcaca	acttaaaat
12721	tttaaaaaat	actgaaaaaa	aaacaaaagg	taacttttgg	ataatgtatg	atactaaaga
12781	gggcatgcac	ctaagtaatc	agactgcttt	tagtaagaaa	ataagttttt	ttttttttta
12841	attttctcag	acaggggtctc	tgtaacccaa	gctggagctg	agtgacacaa	ctcaaggctca
12901	ctgtagcctc	aatcttcagg	gctccaggga	tctctccatc	tgactctctc	tggaagctgg
12961	gagtaggcat	gtgccaccat	gcctggctaa	tttttaattt	ttttgtaga	gaggggctg
13021	ctgtatgttg	cccatgtcgg	tctcaaaact	ctgggctcaa	gctgtactcc	caacttggct
13081	tcccagattt	gggattacag	gtgtgagcca	ccatgtctgg	ctgtctcttc	tttttgtatt
13141	ctaaaaattc	aaggcctaa	tatcaaatcc	ctaaactctc	ataagttttt	acagataaag
13201	actcaataat	aaactccctc	cgaaagttaa	gacaggctca	gggtgagagac	tgttttcaa
13261	gggtttataa	agaaaaaccc	agtgctctgc	agaagaatca	tttttaattt	tttttaattt
13321	attctattta	tggaataagt	tgatcataga	tttgtaaaac	aaaaggtaat	ttctcaagta
13381	tttgggaaat	agaaaaagcc	tcctaccacc	caaccttttg	gtcatctttc	tcttactctt
13441	acaatcatcc	taatccctca	gtacacccct	accatatatc	ataaagggca	ccatataatt
13501	attgcaaaag	cagatatata	tgctgtatct	cttattagac	ttgcaccaga	cagctgtgat
13561	ccactccagg	catgaactcc	aaagctgagg	ccactctgac	aagccccctg	gcactctacag
13621	aagcaaaagg	gttctctctc	cagctggctg	ctcctctcgg	aagcagccct	tactctgggt
13681	taatcgccga	tagagcctct	ctctcaatgg	aggaatgtgg	tagcttaggg	cgaaatggtt
13741	tgctcaactg	cgctctctct	ctgtgttcca	taacttgtat	gggtggggag	gcgaagaagc
13801	tcgtcctctc	tctcttcgaa	tctgtaacgt	aaatccactt	acagccagca	tgccccatct
13861	tgccaggata	atgaagtcta	aaaaacagta	gaaaaagtac	agtttggaac	ctgatttatt
13921	aaaaactaac	ataataatcc	aaccttccat	ggactaaact	tgtatatctc	ttcttaaatg
13981	actttctcgt	gagtttagtc	aaacacatat	gtagtctctc	agttatacaac	caagggaatc
14041	gtaggggaag	ccaggtctta	tttttgaac	taggaatttt	ctattacagg	ctatttgaag
14101	tcagacagag	ttgtaaaaaa	atcagttggt	tggtactctg	gccctccaac	aatattgata
14161	aagtaccacc	agttgtttgt	gtgttttttc	agagggggtg	cggggaggaa	ggcagtgatc
14221	tagttacact	tcacatatgt	acatcatgga	acatcagatt	ccccaagctc	actttctctc
14281	ccatttccact	ctacatggg	ccctactgca	ctcacagtgt	aggttatagg	atgctatgaa
14341	agatttcaacc	aatcgtactc	agcttctagt	aaagcttcta	gcttggcaca	acactgtctg
14401	gtacttataaa	cttgcaagag	tttttagggt	ttcaaggatt	tcaaggtctc	taccataaac
14461	ctctttatcca	aagggaagcc	tctggatagt	aaagggaaag	atttgcctca	aaataaaact
14521	acaacaatca	caattgtgct	tatgcaatat	atcttcaatt	ctaggacatg	ctttttccat
14581	ctccacatgt	caacatttct	gaagtcaaca	tatcttcaaa	tctgaaacat	ctactgtgga
14641	gtcaattttt	tcttcaaaaa	atttatagaa	aatctgtaat	ggatagaact	ttacaagtga
14701	gggaacctat	atgcttacta	attaaaaact	ctttcaacta	gcaaacctga	tgcttcatca
14761	aaggaggaaa	acaagggaat	aacttaacat	taaatcaggt	tgatcatctc	atccccattt
14821	ccatccccac	ccataagctt	caaacatctt	tactgtcagt	attttctctc	ctctgtagaa
14881	tgaccagggt	aaacctaaata	atcttggaaa	tttcttaagt	aaactcattt	agtagatttt
14941	tattctctatc	ttaaaaattt	aaatgcttta	ctccatgat	ctcagttatg	ataaacagac
15001	ggttctcaca	gcattgtatg	ctgtgttccc	tccttttgtc	ccagttgact	ttcttttaaat
15061	ttagttttcca	agctgctctg	ttagcatagg	tggttcaaga	ctcctactag	gctgttcaac
15121	tggttttttgg	atttgaaggt	cttaccatta	gtttgaaaag	gcacatttgt	gaagctctctg
15181	tggaattctct	ttgtgagcgc	tctcttgagt	acgttcaagg	ttagtgtagg	aaaacacact
15241	gaccagtaac	tgtcaatggc	taaaaaccac	gaataggagg	caatgaactc	caaaagtcatg
15301	attgtctcag	gtggaagagt	ggaagagttt	agggcatcatc	catgttagaa	caaaaatagt
15361	ttataactct	ccactggtct	caatcacagg	caatcacatc	tgctccccca	ggcatgaaaa
15421	agaccacagc	ccactcagag	ctgaacagaa	gttccctgga	aaagataaac	gaaaacacac
15481	ttcaacatcc	atataggttc	ttaaaaacaca	gttcaagctta	tccaaaatac	atgggaaaaat
15541	tcttccatag	gaaggaaact	caacaataaa	atctaattag	aggaagaactc	ctcttctctga
15601	agccagttga	aactaatgac	ccatcaggtt	gggtgactag	agtaagcact	accttctctg
15661	ctccactctg	taaaaagcata	ttgactgtga	tgctcattt	gggtgaaaaa	aaagaaaaaa

15721	acaaacaccca	tgtggttaac	ataaaatagaa	cttcgttttc	acagtaacaa	ggctatttag
15781	aacatgcata	ctgataacag	taacacagcc	ctccccaggt	aataaaacat	gcaaaattga
15841	gaataaagtg	atcagctgaa	cacacagga	aacataatca	ataaatctaa	cttttgatga
15901	aaactgtact	atctctaggg	agcccatgaa	aatcacttga	atgaggatga	ctatgcagaa
15961	gaaggtgacc	cagaatacac	catcatctat	aaaaactctt	aggtttccctg	aaaaagcaaa
16021	agaagactaa	gagtaacaaa	cagctttacc	aggttagagt	tggtttctctg	gaacatctga
16081	acacagatat	acatcaaaag	cttaagtgaa	caccagggga	ataatgttgc	aactagaat
16141	gaattctaat	ttctctcat	taactgtcaa	tacaatcttc	cttgaggagt	gtgtctttag
16201	ggttgggtgg	gaggaaccaa	caggggtcaa	gctctgcact	tattacaata	ctattttaga
16261	agaactgga	atgttccaat	ctcttccatt	catagctct	gtgaacttga	gcttcagtt
16321	ctcaaacctc	aaaatgacat	tgaattatat	ctccaagctc	taatacact	gttagctcta
16381	caattgtgac	ttcatgattt	catactgoaa	aataaaaaat	acaagttaaa	ggcatctcaa
16441	gttaagccat	gaacccaatg	taaacagtga	gtccacttaa	cagattttta	cgttcaaa
16501	atgtctgttc	gtcccaactg	acactgtgat	atgtgtattc	ctatgttatc	aggaagagg
16561	aacacaccaa	ctatctgctc	ttcccaaac	ctctctcctg	tgctggaatc	taaaatctg
16621	cacaacttag	gcgaccacag	aacaaaata	tgagactctc	ttcttttaaa	taagtttaa
16681	aaagcaagaa	tagcgtgact	gctgtttaat	actgcactgt	gcaccaact	tcagtcact
16741	tgtaaatgac	cccatacttc	aagacctaca	tatccaggtc	agatgggaag	actgggagt
16801	octagagctac	aggactggac	cttgagactc	catattctct	gaagccaaca	gcataatg
16861	atttaggtct	acttatagat	tctattctct	agtttaacta	tattaacctt	acaaataac
16921	atatgtggct	tcttctgtaa	aggctttcaa	gaattctaa	tattgaaat	actatcgat
16981	catcatgtgt	aagaaaggac	cccttcagag	aataagctca	tgcttccca	cgtaaggta
17041	aaactagcaa	tctaatcaga	ttgtagtgt	cagccaaata	ttcttttaaa	actaatgctt
17101	aaaaatagc	gtaggcctct	tgccgggcca	tggtggtca	cgctgttaat	cccaactat
17161	tggggagccg	aggtgaacac	atcacttgag	gtcaggagt	tgagacacag	cttgccaata
17221	tggaagaaac	ccatctctac	taaaaaata	aaaatgagct	gagcatgtga	gtgtgtgcct
17281	ataatccagc	ctactcaagg	ggctgaggca	ggaaaatcac	ttgaactcgg	gaggtggagg
17341	ttgcagtgaa	ctgagactgc	gcgactgcac	ctcagctcgg	gagacagagt	cgactccgt
17401	ctcagggaaa	aaataatgtg	tgtgtgtgtg	tgtgtgtgtg	tatgtgtgtg	tggtgtatga
17461	tatatagttg	ggccagggtc	agtgggtcaa	gctgtaatc	cccaactctt	gggtgaccca
17521	gggtgggagca	ctgcctgagc	ccagtttaag	accaccctag	gcacagtgta	gactctctgt
17581	ctactcaaaa	taaaaRctta	ccaggcatgg	cagcacacac	ctgttagtcc	agctactgtg
17641	gaggtctagt	taagaggagt	gcttgagcct	tagagtttga	gggtgcagct	agctatgata
17701	gtgtcactgc	agttccagcat	gggtgacaga	aagagatcct	gtctctaaaa	ataaaaaata
17761	tggtgtttga	cccactagca	ttgaggagat	gatcactcta	acggtcctgt	agctatgaga
17821	tagttcttca	acaaacatat	ataaagggtta	tatgctaaaa	ttattttttac	caataggat
17881	agttatgtca	aaaagtttag	agacaaactgc	cttagactac	tttataaagt	tgaaaataa
17941	gagaccttct	tatttctggt	ttctcatctc	tcacatctat	cccttagagat	agtttcaag
18001	atgatggaa	aaagtttcaa	tacatcgctc	aaagttataa	gctagaagag	gaactagtga
18061	cagagccaga	atgcataatg	aaagtttttt	ctgactttaa	taattttatgc	ctattctcct
18121	catgaaagt	acattaaatt	actagaattt	ataacctata	atcccttcat	ctcttaagca
18181	ttaacaaaag	aggggaagcc	cattaaagtag	tagttgggtt	ttgtgtgtct	ttttttttgt
18241	ttttgagaca	gggtcttgtt	cagtcgcccc	ggctggagtg	caatggcacc	atcatagctc
18301	actgcagcct	caacctcccg	ggttcaagca	atcctctagc	ctcagctccc	caagtagtct
18361	gagaccacag	gtgtgacta	caacacctgg	ctaatttttt	tttttttttt	ttcagtagtc
18421	atggggaatc	actatgtttg	ccaggctggg	ctcaaaactg	agctcaagca	atcctccacc
18481	ttgtgtctccc	aaagtgtctg	gattataggc	atgagccacc	atgctcagcc	attaggaagt
18541	attttatcta	gcacatgtg	ggttagagct	tgaattctga	tttcaaaaaa	atccccatg
18601	gaatctgtca	tttcgccaag	cctacacac	actgtctttc	atcagactat	cagtagtctt
18661	tagaaaaaaa	aagctttcag	agagacaagg	gacacatgac	ttgtactcta	ttatctcctg
18721	catctctcat	agagttgaca	gcctcatgaa	atggctattt	ttatgtctaa	cagttcccat
18781	aaatttatct	tagtatttta	agaacattgt	cacagaaaag	aaacaaccaa	actacaactc
18841	tttagtaca	gatgctgttt	ttgtttgttt	ttgacaataa	tcaggaagag	taagttatta
18901	aaacatggca	agaagagtta	acttataagt	taaccagatt	ttaaaaaactg	atagcaattg
18961	aattactttt	taaaaaacac	aaaaaaaccc	cagaaacagg	ccctgtctac	tgccccaggg
19021	tgattaaagt	cgtgaggac	acctggggtt	tgccagcgga	tggaagtgga	gggtgaggt
19081	acaacagaag	accagaaaac	aagtaccagg	ttccccgact	cagctggcat	cagagctttt
19141	acctcttatt	ttcttttaac	tccattattt	ctaaataaca	ctaccactat	actacaacaa
19201	atagatacaa	agctacctct	aaagacattc	tatatttcat	ctactgtgcc	tttgaggaga
19261	ggaggacttt	tactagctct	ggtgctgtct	gaaagtgtca	caccatcaat	cagattttcc
19321	agaactccac	aggagaagca	ctgacatgca	accaccctcc	agagatgcata	tccaattagg
19381	cagatttcca	acagtcactc	gaacacacaa	aaactctgtt	acatatgcgc	tccacaggtg
19441	agtaaaagaa	gtcactgacg	agatgaggaa	ccccagcact	agtccaacac	agagcatgca
19501	gatcgagagg	attccaaatc	gccaccacac	agctaccaag	aaactctccac	cgacgcttcc

19561 agtgacagct gtcagaatca gattcaactgt acaaaagagag aggtgaact cgtggtaac
19621 cggcagact tccaaccgag gctctagac atacaatacc gctaccagtg cgcagctcag
19681 cttctctag gtcaggcaact gagggtcttY cagtggtgccc atgaaggcag catagagcaa
19741 cccctctag ggagggtctt gggctgagac aaggaalata ctgaggagc taaccctcaY
19801 agggccagtg acaggcttgg ctgtgcaata aattgggatt gagaaaggaaa agttgacaaag
19861 cttctgtta aacataatgg ggtaggggag tatctgtttt caggtaaaga aacctccaac
19921 atttttctgt ttgtacatct catatcttca acatatcaca tagccagaac accaggtctat
19981 aacaatattg gccctctctt tccctaagta coaaggaaat ccacatacag attaaaaaac
20041 cagaagctgt gaatgagtag ttcagaaaaag gttgtcagag actagatgga ccatattttt
20101 acataagtaa taggtctgtca cacacaatai agggccaaact gtagaaggaa aggtgaagta
20161 aaacacagca cctgccctctt tgaataacca gttaaaactat ataattttca cttaccatca
20221 tacttgatg gtgtcagctct tgtaatcagt atataaaaga ataatcccat gatgataag
20281 cctatgaaga ataattYraa gtggccaaacc accaggaaag gaaaaaagca aaacacaaaa
20341 agctatggtt aagcacctct tttacagtg aatttaata tttactctac aaatcatgac
20401 ttaagctctct ggaatttaaca tgtttacagca ttcagaagaa aaaaaaaatc taaagtcacg
20461 aagaaatcaa gcagccatgg gagtaattat tatataacag atacaatat aaagtccaa
20521 ctacagtaga cactttctctg tatataagtc agtttccatt gttaaattgt tattttgttc
20581 tttctatgct cattttaata tttttatgtt tttctgttaa aataagataa gacattttt
20641 acgtgtatgt gatcacgaat tacacattaa attttaatag tttgttaagc agaaataaaa
20701 tgtaaaaagt aaaaagaacc caaactgtcc caaagcaaaa tataactttt atctctctt
20761 ttaattcact ctatcctgtt tagaatgac atcatattacc ttcactcatt aaacacctg
20821 gtaatgcag tatcacagata actttgttcc cacaaaataa cctaatctgt atcatggagg
20881 ccatagctct acctattttc ttaaatctct cacagctgtt cttgttaata tttgtactgt
20941 attctaatct caataaatat aaaaatgMta atatcaaac accaaaaaaa cctagagtga
21001 ggcattggtg cttgtgccgt tagtctcagc tacttgggag gctgagggtg gaggatacct
21061 tggagctagg aagtgaaggc tgtagtgoac tataattgta tttgtgaaga gccactacac
21121 tgcactccag cctaggagac acagttagac tccattctct taaaagagaa aaaaaaacg
21181 tagtgaatc tttgagacca cagcttagca actaagtctc tttgtatcag aaatcatgac
21241 atgttgagat attatatatt ctttacaata ctctcttccc aatttaataaa aaaagtgtca
21301 taatagcagg aaggagttgt ctatttcaaa caatttaaca taagtttaaa gactttatt
21361 agatcaggat tacattttgag ctaccttgat gctcagttat acaactgttt gattaaaaaa
21421 ggaatttttta ataatcgaaa acatcactta tataaaaagc agcccaata atctcaataa
21481 attcagtgag catgattttac atccagtttc ttcagttgtt ttcactgatg ttgactgtcg
21541 cacttccctt gccactgaa tctctcagga ctgagcagac ctgtcaaac actgtacac
21601 tgcagcccca tagcctgatt catgtgaaac caagtcaact gccatcacca atatccacg
21661 atatttttca aaagacaacc aacttctttt tttttttct tttagactctg tcttactctg
21721 tcatccaggc tggagtgagc tggcacaact trgtctcact gcaacctctg cctcccgggc
21781 tcaagcaact ctccacctc agcctcccca tgtagctggg ttagctcagc cacaccaatca
21841 agcctggcta attttttgta ttttttgag aaacagggtc tcacogagtt acccagggtg
21901 gtctcaaaact cctgagctca agcagctctgc ctgctcagc ctcccaagt gctgggatta
21961 caggtgtgag ccaccacaca cagcccaact ttgtctttaa ttttttgta gagaaaggat
22021 ctcaactagt tgctagact tgtttcaaac tctcggtctc ctctagctct ctaacctg
22081 ctcccaagat tgcctggatt acaggcatga gccactgagc coggccctga caaccaagt
22141 ctaactctgg ttatctttag tcaactaaaca gatctagcag gatctgaaga tctactgta
22201 attgtatctc cctgttaata agagaagggtt ataactcgct gtgaggagaa tagaaaaaaa
22261 taaaaaatgca ttgcaactca gttagactcac gactacaaga cccagctac aactctatga
22321 agatagtaga gataagttag tgcttactct cctcacaggt catctccag caacactgaa
22381 aagactgcga aatggaata gcaactacat tggacataat agctgttttc cagatctgt
22441 gtccaaaagaa acaaatgaag aaaccaaaga gggcaaaaag agtgaagaac actttggaag
22501 acactctctc taaaaaataa tgaaaattta tgggataata ctccaaact gacttttgt
22561 attaatattt atagaacgaa ccttcatcaa aaaaagctaa tgaactgaca taaattatcc
22621 catcacgtga ttatcaaat ttatcttttaa aatactcttg aatacttca ccacccatca
22681 tatgtagacga atcaataatca acaactgtgt ctatacaaaa aagagaatcc agacaataa
22741 tgcctgtgta tagcggtagt gaacacttcc tatgacatat tctgtggtaa agtgaggaga
22801 aacagagaga gagtgttctt aaaaacaaata atccgttttc ttaacaaat aaactgtatg
22861 gacaaaaggt gcaggaggt gggggggagg tgggtggggg aggtgagggg aagctacagg
22921 taaaagagca tttaaaagac acatcaacta aatactgtgt atagatctca tctgcaacct
22981 gattcaaaaca aactgtggaa acaaatgag atgactccag aagcaaaactc ggtctgttaa
23041 ttgtgtgcta ttaaaaagaa attttcaata ttttaaaagt aaaaattagt taagtgcacg
23101 ggtgtctat tttttaagtg ctatatctta aggtatcac actgaaatgt tgtgagtaa
23161 catgattgacc agggatttgc tttaaaataa tccagtgata aaggaggggt atgaggttga
23221 gaggcagaga agcaggtata gcacagctga aacaaagata tggcacaata ctaaaattg
23281 taaggctgga taattgtgtt atggaggctc gctgaagttt ttcacaaata aaaaatttaa
23341 ccgaactca caatttaaaa acatgtatta cacttaataa ttatttttat taatttttat

23401	tacacataca	caaattacta	ccagtccttat	caattataca	tattgccttg	gagaaaaatg
23461	tgtcattaca	accaaagtgt	aagcacggta	tcaattagtt	tactgaagtg	aaaaatattg
23521	tcactaaaat	catcacccag	catcatgaga	taaacccaag	aaaaataacc	ctgacaactaa
23581	acccaaaata	agacatctat	acaattctcat	gctaaaaata	acacaggctg	ccacaaatca
23641	aaatcacata	aagtcctaa	agtgtctaga	atgggagaac	catgaaagac	tacagtaaaa
23701	taagaaaac	agacagaaga	gcgaattcca	caagggatata	aatacctaaa	ctcagctctt
23761	tctgagaata	aggagcccat	aagaagaatt	aaacaaacag	cggtgtcagt	ggctcacgac
23821	tataatccca	gcactttggg	aggccgaggt	ggcgagatca	gggtgttgag	ctgtgtttgag
23881	atgagccttg	gcacataggt	gagaaccccg	tctactaaaa	atacaaaaat	ttagcgaggt
23941	tgtgtgcaca	tgctgttagt	ccagctacct	ccagaggtgt	aggcaggaga	attgtctgaa
24001	cccaggaagc	agaggttgca	gtgagccaa	gatcaggcca	ttgcactcca	gcctggagtg
24061	cgagcgagca	ctccatctca	aaaacaaaaa	acgaaaaaca	aaaaaacacc	acccagaaga
24121	attaacccga	agataaaaac	accacattgt	ttatttcaga	cctgtggttg	actaatattc
24181	tagaatttta	tttttacacc	aaaaccatta	aatttgtctg	gatgatitct	tcaaatgacc
24241	atcccggtga	cttttcagca	agataataca	ccaagacagc	aaagctcttc	accaaatata
24301	aaaactaaac	acagtttagat	actcttccat	aaacttagag	gggaactaaa	tcocctatca
24361	caataccagg	aaggaggtag	agaaaaaatt	atacaagtaa	aatgagagggt	tatggaagtg
24421	tccatacaaa	gcgaattcta	cgataaaaaa	agacataaaa	caggctgacc	ctctgggaag
24481	atgaaaggag	tgaatagact	aggacaggag	caagttgaag	gttaaagtg	cttaatactat
24541	gacagctatt	tgaaagggtta	tgaaccacag	ggcaaggttt	tagttaggag	gcnatttgga
24601	tatagctccc	ctactccaa	ctcaggaaaa	gcactctaat	actactcaat	cttgaaatag
24661	agagtataca	tagaactaac	atcactagga	aggttcaact	ttggcgggga	agaaacaact
24721	gatcagagct	ctctgtctcc	taagccctaac	agagagagac	aggcctgggg	gcctttctat
24781	ttgagagggt	cttttgcatg	aggggttcaa	gggaaagaaa	tggaggagacc	cgacatcact
24841	tcagtccatt	ggttaaaaa	atatggccag	gcgcagtggt	tcacgccttg	aatccccagca
24901	ctctgtggag	ctgaggcggt	tggtatcaatt	gaggccaggga	agctccagacc	agctctggcca
24961	atatgtgtga	accccatctc	aactaaaaat	acaaaaatta	gtcaggcaag	gtggccgcca
25021	cctgtaggac	tttgggaggt	cgaggcggtg	atatacctgt	aggttcaggag	ctttgcagcca
25081	acctggcccaa	catgttgtgt	caggtctcta	ctaaaaatac	aaaaattagc	aggtgttgtgt
25141	ggtacaccca	tataatccca	gatactctgt	aggtcgaggt	cttggaactta	cttgaacctg
25201	agaggcgagt	gttgagtgca	gctgagatca	tgccactgca	ctccagcctgt	gggtgacaaga
25261	gtgagactgt	caaaaaaaaa	aaaaaaaaaa	aaaaaaaaaa	aaatatatat	atatatatat
25321	atatataata	acacacacaa	acatatctgc	caaggtctct	tgcatttgat	ccctcaagga
25381	tcttgagttg	aagattgtct	tgagtggaag	actgagttgt	gacaaagagc	ctttgtctgt
25441	ccaaattttct	ttttcgtttt	tttttttttt	tttttttttt	tggagacagag	ctctcgctctg
25501	cgccccagcg	lagagtgtag	tggcgcgatc	ttggctcact	ggcagctccg	ctcccgccac
25561	tcacacattct	ctcctgcctc	tgctcagccc	tcctctagtag	ctgggaactac	aggcgcccaa
25621	caccatgcgc	ggctaagtgt	ttgtattgtt	tttttttttt	gtagagacgg	ggtttcaactg
25681	tgtagccgat	gatgttctgc	atctcctgac	ctcgtgatcc	ggcccgctgt	gcctcccaaa
25741	gtgctgggat	gacgagcggt	agccacccga	cccgccctgc	ttggccaaat	ttcagttccc
25801	tgaacccctc	cctaggccca	tctgcgcaca	tccttgtaaa	atccagcttc	agcaagaactc
25861	ctgataagct	agcttagcaa	gaatccctta	tgcttgatat	caaatcaccc	tgcatactca
25921	atcaggttct	tcactctcca	ccatccatcc	cccaggtgat	gtctgacacc	ctcaggtcttc
25981	ctctagtaaa	aatcccttta	ggtcagtgca	gacagaaatg	ccctctgccc	ctcagttctc
26041	ctcttagaata	ttttccatcc	actgactccc	acccactctg	ttggctacaa	atttccacgt
26101	gccccattgt	tactcaagag	ttggggccaa	ctctctctcc	ccactgcmaa	atttccatcc
26161	agtggtccct	atgcctattg	taatactacc	ccccaccctg	aataaaaatt	tcctcactgt
26221	ctcttaacta	gtgtcactga	tttttttttt	ctttaacagc	gggatactgc	ctagacagag
26281	gtagagaaccc	ataatctctc	gcttgaacac	aaactaataga	ctttctctct	tagagcaaa
26341	ctccttgaaa	gagctacttg	aactttctcc	agttttctgt	ctccacgctc	ctatttgttc
26401	tgaattatgt	tccagtcagg	cgctaccccc	acctctccac	tgcacagcgt	ctcacttgta
26461	ccaatggcct	ccaattgtgt	aaatccaaag	ctttaacagc	gggatactgc	gtctctttct
26521	atcagacaga	ctgtccaaa	ttgattactc	ctctctccat	gcaatctagt	ctccctctgg
26581	ctcaacatca	catcaatttc	cttttaacaa	ctctggctga	gcttctcagc	catcttgtag
26641	gatcttcaat	ttcctaagct	ctaaaacattg	gagcacccag	gggtcagctc	ccagaccctc
26701	ccctttttct	taagtcaaac	tgaattgtgc	tgctctccc	acctccccct	tccactctgc
26761	ctccccctta	tagataaac	catttagact	gagggcctaa	atccacata	ttcagttctga
26821	agatttccaa	gatgctctcc	agctaagact	ttctctgtaa	tctcagcgtt	atatataata
26881	tcaaacatca	gttlaaacatc	tgcacataaa	tgtcctaag	gggtctctaaa	ctgaacttgt
26941	ccaaaactga	actcttaagt	tatcctccca	accttctcct	ccaaactcgc	ttcccaactc
27001	agcggaattgt	cgctctctct	ctccagttgc	ctcaggccaa	aaactacaa	ctctttacac
27061	ctctttctct	caaacccctc	atccaactca	tcagtaactc	gtaactccaa	catcatcagc
27121	caactcatgag	ccaaactctg	aacaatttga	acatgaaaa	aaatgatagt	aaagatttcc
27181	atccagagaa	taaaatagga	atctacaggt	ccaggtttaa	aatataatca	tggataaata

27241	aatggcaggg	gagggaanaac	ccttctcag	agcaggatgg	caacaaataa	acatagaagg
27301	agtgatacat	tagagaaaca	ccatttggc	accacctagt	aaataaatgt	tcaggcaagg
27361	atccatgat	gccaaacta	gcaagtgaga	gtttgagaat	gggatattta	ckggaactgg
27421	agatactcc	cctaaagaca	cttaaaaaa	ggaacagtta	cttcagttgg	gaactttggc
27481	aagacaccaa	taaccaagtc	atcaaggatg	tcaccaccag	taaaagagca	aaaaaagcat
27541	gagacaagtg	ttccttgaga	aggacacatc	actctcttgg	tattctcgcc	ttctaaaaaa
27601	aatcaaaatt	taattatgag	gaataatcaa	agtaacccaa	attgaaagca	ttctaaaaaa
27661	aaataaattt	ctgtacttct	caaaaaagtc	aaggtcttaa	aaagattaaa	ttctaaaaaa
27721	aaaaaaagac	ttgttccaga	ttatacaact	ataaagagac	acatcggtta	aatgcaatgt
27781	gggatctcaa	actgaaccc	ggacagggga	aaaaaaaatt	gtgtgacggg	gagccggggg
27841	ctcactctgt	tgcccaggct	ggagtgagga	agcacaatca	cggtctactg	gagactccac
27901	ctctctctgt	tcaggtgata	ttcccacctt	agactaccaa	gtagctggaa	cgacaggcat
27961	gtgccaccgc	accagcttaa	tttttgtatt	ttttgttaga	acggagtctt	gccatgttgc
28021	ccaggtctgt	cttgaaactc	ggagctcaag	caatccggcc	acttgacctt	cccaatgtgc
28081	tggtattaca	ggcgtgagac	actgcaccca	gccaaaaata	attttttttt	acttataaag
28141	gacatacttg	taaccaactag	taaatgtgct	aacgtgtgaa	ttaaataact	actgtatagc
28201	tggttatgct	gtaattttga	taattttact	atggttatgt	aaagaaataa	atacagaagt
28261	atttaattgc	aaaaaggtct	ttatgtctgc	aacttactct	catatggttt	agagaagaat
28321	atttgttgtg	tgtgtgtgtg	tgtgtgtgtg	tgtgtgtgtg	cgatctcata	catcacaggt
28381	cgccctcaga	acaaacacgg	tttgaactgc	atgggtccac	ttacatacac	actctttttt
28441	agtaaaaagt	acactgagtg	tgcttgcctc	tcctacctcc	ccttccacct	ctgcaccccc
28501	tgagacagca	acaaacaccc	ctcctcttcc	tcctcctgtt	cttatctaat	gtgaagacga
28561	tgaagatgaa	ggcctttatg	atgatccact	ctcacttaat	gaatagttaa	tatactttct
28621	ctctcttatg	atttctttaa	tgacattttac	ttttctctag	ctcattttat	tgttaagata
28681	tagtctatga	taaaataaac	atacaaaata	tgtgttaact	aaagtatttt	gtaacttaata
28741	aggcttctct	gtgcaacact	aggctaatag	tagttaattt	tggggaagtt	ataagttatc
28801	acgggtgggg	tatgtgtggt	tatgcctgta	atccagcac	tttggggagg	caaaggcggt
28861	ggatcacttg	agggcaggag	tttgagacca	gcctggccaa	catgtggaac	cccagctact
28921	actaaaaata	caaaaaattag	ccagggatgg	tggtgtgtgc	ttgttaattc	cagctactca
28981	ggagctgtgag	cgaggagaat	cacttgaacc	tgggagcgag	aggttgcagt	ggccgaagat
29041	actgccactg	cactccaagt	tgggcaacag	gagtgaaact	atgtctcaga	aaaaaagaaa
29101	agagttatac	acaggttttc	aactgcacag	gagtcggctc	ccccacacac	ctctacactg
29161	tcaaaaggtca	attgcatata	gagagagaag	aaagagagaat	gaaaagcaaa	tagcttaata
29221	acttaagctt	tgaggaaact	ggatgaagat	tatacagaaa	ttcttttatg	tattttttga
29281	tattttctgc	gaaacacttt	caagataaac	tgctaataac	acacacacac	acacacaccc
29341	ctcacaaacac	cttactctgc	cccattctac	ctaacttagt	tatctctccc	ctgtattact
29401	tcagttggct	cctaaccaat	ctgcattttc	gctcttgctc	ccgtcttttt	tcaaaaaagc
29461	gccagtgtgc	tcattaaaaa	atcttagatc	atggcaccat	ttgtcaaaaa	ctcttcacga
29521	ctcttccctt	tcattcagaa	tagaagcaaa	atccctactc	acagttcaca	agggccacac
29581	caggctgcct	gcctctgcac	ctctcacctc	acctctactt	ctcttgccct	ctcgtttctc
29641	ctgcaacccc	gtctatcact	gcagttcttc	acacccctag	acagggctcc	aggtgcaggc
29701	ctctactact	gtgctccct	ctgcctggta	catgccacac	ccaggttagct	cgctgtgtga
29761	tcaaggactg	ccctggccat	gctatctatc	tcaaaatttc	ctcttactgt	catgtgttat
29821	ctccctttcc	tactttattt	atctctctag	cactcaccaat	ttctctacgt	attcatattt
29881	ttataaaatt	atcttgtctg	tctcatctcc	ctcactataa	tataagctcc	agaacaaagta
29941	tttttgtctg	ttgttttact	gctatgtccc	agtgcttaga	acagagagtg	gaataacgta
30001	ggggctcaaa	aggtatgtgt	tgaatgaatg	agaactagac	aaaagcatta	ttgtcgaagg
30061	caaaaggaag	tttgaanaag	tttaaaagca	catactcacc	taggggaagg	caactaacctc
30121	ctcctgctcc	aaagctgcaa	gcgtatgtgt	gagcaggaat	gtaggcagca	gatgtattta
30181	gaacgggtgc	caaacacagt	acattgtata	tgacaccttg	tcocgggaag	gaggaagaag
30241	aaacacttgt	cttatcatata	gctgttaggg	taaccacctg	aaagcaaaaa	taacacatga
30301	tgtgtgttta	gggatttttt	ttaagctatt	ctgtgtataa	cccaccctacg	tgtcaattta
30361	aatggtatata	tcataaaatta	taattctctg	ttaaaaataa	caccacaccca	tttaatatgt
30421	tcaaaagccc	actctacatg	gtttgtatgt	gctaagttaa	acaagctcca	accacaaatg
30481	ctctccaaag	ttttttttta	tagtttaagg	aaaaatcgat	ataaaccaatg	tataaatgct
30541	agttacgacc	atgaaaaaag	ttttttcacc	ctctgcagac	ttataaacatg	acacagctac
30601	gtatttctct	tgctcactgc	cttctctcca	gtgtacagcc	caatgcctaa	cacagagagt
30661	gtgtccaatt	ttagccacag	actaaacgag	acagcttggt	aacatctcgc	agacagttacc
30721	caacttattt	tgcttaacag	tctctgaaac	aagaaaataa	tacatgaggt	tttgttatgt
30781	agtagtagtt	cttttttttt	tttttttttt	tttgaacgg	agttctcgct	tgctcccgac
30841	gcggagctgc	ggaactgcagt	ggcgcaactc	cggtctcact	caagctccgc	ttcccggggt
30901	caagccactc	tcctgcctca	gcctccccag	tagctgggac	tacagctccg	ctgccacact
30961	ccgggtcaat	tttttgtatt	tttagtagag	acgggggttc	acctgtgtag	caggaagtgt
31021	ctcgatctcc	tgacctcatg	atcacccgcg	ctcgctccc	caaaagtctg	ggagtacagc

31081	cgtgagccac	cgcgcccgcc	cagtagttct	taacttcaga	aattccattg	taactataaa
31141	aatagaagaa	acaaatgctt	tadgaacca	taatacctat	gttctactaa	attttgagaa
31201	aaaaataact	agggttttag	gagaaaatta	ttagtagccc	tcctactaca	gactaccatt
31261	gtttataaat	actgcttagc	cttcaataac	ttgtctctct	cggaaggctc	gtttaaagtt
31321	acacaaactg	ttaataattac	tcocaaagta	caatgaaaag	tttaaacacg	acagaaaagt
31381	tgaaatgaata	acacaaataa	tagactataat	actgcctaat	atacacaatt	gttaataatt
31441	ctattttgct	tctctatcac	tagtttctcc	cnaactctac	tttttctctg	ctgaaaagta
31501	gtcacagata	tctagacatt	tgacttctaa	acacttcaat	gttctgctcc	tacacagctg
31561	tttctcaaat	ttaacgtgct	ttgtaatcgc	agggggcagt	gttgttaaaa	ggtagatctg
31621	attcttaggt	cagctgtggt	ctagcaattg	tgctttctca	gaagctctgc	agatgtatgc
31681	aatgtcagta	aaacacctac	cacactctga	gtaacaaagt	actcagtaca	gatgtaagtc
31741	taagtacaga	cttacaaggg	ggttacatca	ggataaaccc	acgtaaaagt	caaatgcgat
31801	ttaatgtacc	tcacccacca	gattattatg	ctttacccta	gcctacctta	aatgtgtcta
31861	gaacacttat	attagattca	gttgaacaaa	atcatctaac	acaaaggcta	ttttattttt
31921	tatttattat	tattattatt	attgagacgg	aatgtcgtct	tggtgccacg	ctgtggagtgc
31981	agtgggcaga	tcactctggc	tcactgcaag	ctctgctctc	cagggtctgc	ctctccatct
32041	gcctcagcct	cctgagttag	tgggactaca	ggcgcccgcc	acacacacta	cgccccagcta
32101	atttttttga	tttttagtag	agacagggtt	tcacctgttt	agccaggagt	ctctcgatct
32161	cctgaccttg	tgatctctcc	gccttggcct	cccaaagtgc	tgggagtaca	ggcgtgagcc
32221	accattctcg	gccacaaaat	gctattttat	aataaagtgt	ttaattgtct	atttaattta
32281	ttgaatacag	tacactgtaa	aatgttagct	attaccctgt	tggttaagca	ggagcagatg
32341	ctcactgcca	ctgctcagcg	ttgcaagagt	atcatactgt	atattgtctg	ctcaggaagt
32401	tatcaaaact	aaaaatcaaa	gatcataagt	caaaccatca	taagtcaagg	accactctga
32461	taagaataatt	ctctctata	acaaactctt	agctcaccta	cgaaataagt	aaatytctta
32521	ttagctataa	aatgatccag	attttcccta	atgtctttttt	tttttttttt	ttttttttgg
32581	acagggtctc	actctgtctc	ccaggctaca	gtgcagtggc	gtgactctag	cttactgagc
32641	ctctctccag	gctcaagcga	tcctccccc	tcagcctctc	caatattctg	gactcacagg
32701	acgcaccacc	acgcttgga	ctttttggtt	tttttagtgt	ctgtgagctt	tttgccagct
32761	gccagggctg	gtctcaaat	cctgagctca	agtgtatctc	cctctctagc	ctcccaaggt
32821	ctgtggatta	gtggtgtgtg	ccactgtctc	tggtctctaa	catctttttt	agctttttct
32881	ctctctacc	aaagatcaaac	caagtctcat	gcatggtgtt	tacggccaat	ctcagtaagt
32941	cagagcattc	gtctccacct	ttctctccca	gctaccaaat	gtctgtaga	gaatgaatg
33001	cagtattctc	ataaagtatt	ccacaatttt	gtgtctaat	tattctttta	tcctccaat
33061	ttctctga	tagggaatcc	aatggactaa	attcaggtta	gacatttttt	gtcagaatat
33121	accacaggtg	atggtgataa	tgctcagact	cactacatRa	gcacgtta	caggctgatc
33181	cattttcagt	gaactagctt	tgccacctct	ggaatagggt	gagactataa	ggctacatcca
33241	tgttaaagt	gttctctctt	tacaatgaga	atlaacttaa	ggagtctact	tagcacacct
33301	agaaataatc	tgttcccca	aaacctttca	cctgctgac	cttgcctgaa	ttatttcaatt
33361	gagggtcgca	aaatgggtat	catctaattc	tattatttct	tctacatata	ttagccagca
33421	ctctctcgca	aagagctttc	tctcaccaac	tggggaagaa	ttagtctctc	ctttaaaag
33481	cgaaactcaa	ttcttttate	aatttttaga	gtaatttgct	gtaaggactc	ctctcaggtg
33541	tggaactgca	ggaaggtttc	attgttattt	attcaaggct	gacaaccaca	gctctttttct
33601	tgattctgca	actgcccaca	atttggcagt	gcaagccctc	tcaaacagcc	ctctgaaatct
33661	ctttcgata	cctcatttag	tctttgaggg	tttccccc	ttctgagcga	aaatatttct
33721	ttaactaaaa	gtttaaaagt	aacctacaaa	aagctcatta	tacaagtctc	tatatctctc
33781	ataaaaaata	taatagtatt	aaaaatgtca	atctaataaa	gcaacttaata	gttatttacta
33841	acagagcttc	ttattttgtg	ataaagcaca	atgacgctgt	ctagaaaaac	tcacaccatc
33901	aaccatatga	ctgaagatgc	acctcatata	ttatattgct	gttaggcact	gatataagaa
33961	taagataaaa	atcctgcctc	ccacaaaaat	gatctattgc	agtgagctct	aaaaatttaat
34021	atgcataaaa	atcacctgag	ggctggagtg	caacagcacc	atgctggctc	attgcaccta
34081	ccacctctcg	gggtcaaaac	attctcctgc	ctcagcctcc	caagtagctg	gtatttcaggt
34141	catctaccaa	catgctctgc	taatttttgt	atttttagta	gagacagcgt	ttcacagtta
34201	tggaagagct	gggtcacaac	tcctgacctc	aggtgatcca	cccgctctcg	ctctccaaag
34261	tgctggaggt	acaggcgtga	gcactgcac	ccagctagaa	tcactgagaa	gatatagaat
34321	gcgaattctg	attcagtagt	tcagggtgtg	agtttaagaa	ttcaaatctt	cacaactctc
34381	ccagtgtatg	caatgccact	ggctcatggg	ccacactgtg	ggagcaaggt	tttaactgga
34441	tcaaaaagct	cagagaagaa	ttttactcca	agttagactt	tgagtttggt	ttcaaatatg
34501	aatgccactg	gttatctctt	taagcagttt	ctcatatttt	ttttctcaaa	acaatgtaat
34561	ctctcatact	gccaaaaatg	ttaccaggt	agtcgtggga	gtacacagaa	aatatccccc
34621	tccttctcct	caactttcag	attaccaagt	cacagatgtg	gctgcaacta	aaacgttaaca
34681	taatttatcat	gcattctctc	aaaggccaac	atcattttga	gatataccca	cgaaagactg
34741	ccaagatgat	acaaaaagag	ataaaagtgc	gtgtccccc	aaaaggagaa	gaagataaac
34801	aaataagact	ctggtagcct	ctccctctct	caactctgac	aaaggtgtca	caatccagcc
34861	acacaaatca	tgggactcag	attttacctg	agagactcgg	ccttccactg	gggcacactg

34921 accatccctct gcagatgctt cagcaacatc tctcagtgga ggtcattctc aggcagaaaa
34981 tactgataga ctcatgactg caacctccac ctggagctct ggtctgtccc aggtccacat
35041 ggtgggggat ctacacccctc aaagtcaccc aacaaaatga aacatcagtc tctgtgtgtg
35101 caggtcccat ctataaaaa gacccaaagag cctcaatttc aaacatctgc aaaaataaatt
35161 acaatttaac agttcaacct ttgcagtggt acaaatgggt attctgaaga aaagtataatg
35221 actgtttatgg actgaatggt ttgtctctcc caaaattaat atgctgagac cctcaactcc
35281 agtgtgatga tattagaggg tgggaccttt ggaagttaat taggtcatala aggtgaagcc
35341 ctatgaatc gaattactgc cctcatataa gagatccagc agagccctct agcctctctt
35401 ctgtcatctg ggagacacaa gacaaaccag cagtctgtaa cctgcagacg aacctccacc
35461 agaatccac ttgtgtggca tctgtatcgt ggaactctgc agtcgagac catgagaagc
35521 aagttataaa gccaccaggt ctgtgtact ttgttacagc accacaaact aagacaaaa
35581 ccaacaggca aagtgatta cataactgga ttctgatc agtcaggccat agagtaaaat
35641 actacagctc tggggaggca aaaaaatttt tttaaaaaa aagctgggtg cagtggctta
35701 tgcctgtaat cccagactt tgggaggctg aagcaggagg attgtctgag ctaagaagtt
35761 caagaccagc caggccaaca tagaaagacc tcatctttga tttaaaaaa aaaaaaaga
35821 gaagaatata caagatttgt caattgttag caagtgggt aacctccata acccttagg
35881 aaactttctg taaaataaca gcaattgcc tttctcataa aattgttagg attaaattgc
35941 aaagaactca acaacactca gacccacgga caatgcacag aggttctca aggttctca
36001 aatgttagct actcatctca ttaagcaga aactgaggca cagctttaat taaaacaattg
36061 aaactttata aaaaataaag gatgtttcag ggtactgag tgcattctat attttaaagt
36121 ctttataaaa accaggtatg atttccaag tcaataaga ctcaactgtg gctcttccaa
36181 taaaaagagt tcacaagaat gaagtggaaa aatatcaact gtaattgaca gctccaaatt
36241 taatagtga gtttaatta catcaattcc ccccaaaat taactcaatt tactacaga
36301 gaattaggga cttgagagag ctggctcaat ctacacact ctactcaatt atttctaaa
36361 agcatttctc agacatggga gacagacgta cctgcatag cctaggtttg ctctgggcaa
36421 ctctgatgct gtttcaaaga aattatact caagttaagt ttgtctctca tacttaact
36481 gaactcccaa ttacagctcc cagggacagc atctggataa gaaataggga agttactaaa
36541 caaactaggt tcatatattc aaactcact cagaaaaaca gaactctac tacttctaca
36601 aaaaactccc aaatagaagt taaaataaaa cgttagactt ctatgtctat actatttcat
36661 gaactgaggt taaagagggt agagttaact gaggttcagg cctcagtagc agggcagagc
36721 caaaattttc tgtaaaaggc ctaatagat tttaggcttt gtcagttcga tgcctctgga
36781 agtaactacc caccacgct ctgtgtggcag gaaagcagc acaggccaata cgtcagcaga
36841 gaggcagctg ttgtttcag taaaacttta tgcataaaaa caggcagcaga gctggattctg
36901 gctcatggtc ttgtagtttc caacctttc tctagagtta caactcagc cttattgttg
36961 actactctct gccaggtttg gaatgcagat acacagtggt tatatatgtt tcaatatcat
37021 gttggaactc tctccaaaat tctatgctct tcttctatc tctctctttt ctgcaactat
37081 ggtactttac cacaccacaa atggttatga tgcacagctga agctcctctg tctcaattt
37141 tcttttgtac actgtccttg acgcaagcct taattaaagt tatgtttctt tatcaggtaa
37201 tttagctttt gttatagcta acatttctca tcttctact cagtaattctt ctggtattct
37261 cctcctttcc taactagcgc tccctctctc tccctttacc cccaagcttt aagagtatgc
37321 taaaattttt caactcactc tcaactctcc atccctgttg tgcactcaaa gtgacttca
37381 taaaggggcca ccgctgactc ctactgtctc tagcatacgg acactttcag ctagtctctc
37441 tgacctctcc gtactagtta acttaccact tcttttttt ttttttttt tttagaacct
37501 agtctgtctc tatcacccaa cgtggagtgga agtggcgcaa tctcagctca ctagcagcc
37561 tgcccacagg gttcaagcaa ttctcctgcc cgagcctcY gagtacactg gatgtaggc
37621 accgtgccacc aacactgaet aatttttgta ttttttagta agatagagtt tgcgcattg
37681 gcccaggtga gtctcaaat ccgattata taaattatc ctgcaacctg ctctggcctc
37741 ccaaagtgtc gggattacag gtgtgagcca cgtgtcccg agtctccct cttcaatgca
37801 gaaactccag gctttcttcc taacttctg actactccac agtgtgtgct ttttctctca
37861 gggattatca tccaccagaa acagtgcctg acacatagta ggcactagta agtatttgc
37921 tatgattttt ttgcctaaa tatccaagta catgctcga taactctact cagtgccaca
37981 gcttttata acagctattt agttgggaa ctactctcacc cctccactca caggccctta
38041 tttaaatatt gtaacttgca tttccatcca aatgtctcac aagcatctca gactgtgtc
38101 caaactgaag tcatcacata ctccgtccca cctccacaaa cactcagac caacataacc
38161 ctgttctcat acctattttg tatcactatt tctctctgc ccaaataggga aacctattgt
38221 tctctcttat actccataac caccagtcac ttgttgttat caattccacc ttctggatat
38281 ctcaaatctgt gctctctctc tccatcccg ctctcatggc ctctgttcag atctcatatt
38341 ctcaattttg gtgagtcct gaactctgct cgcctctccc aatcatccac cactactat
38401 ctgagtgatg ctttctaaga gatgaaatgg actacacat tctctgtgt taaaaactca
38461 ttgctctcta gaacctacct cgaataaaa tgccattctc aaattgtgat caaggccctc
38521 cacactccag ctgcagccac actcacaac tgccattctc caactattt ttcttcagta
38581 cgtcagctgt actgcttctg cttagaaca ataaactgtt aaggactagt cttctcagta
38641 actattactg ctttccaaag ggctccctaa atgttttctg aatgaataaa catgttctg
38701 tagtcaactc cgtgactttc agtcatctg tactataaaa atttctgag

38761	tataggtctgg	gdcagctggc	tcattgctgc	aatcccagca	ctttggggagg	ctgaggcggg
38821	cgaaatcaaa	ggctcaggag	tcgagacacg	ctgggccaac	atggtgaana	ggcattctcta
38881	ctaaaaatac	aaaaaattag	ctgggctctg	tgggggcgcg	ctgtaatccc	agcattcttgg
38941	gaggctgagg	cgaggaatc	gcttgaacct	ggagggcaga	ggttgcagtg	agcgcagatc
39001	gcgcactac	actcagagct	gggtgacagt	gcgagactcc	atctcaaaaa	aaaaaatcc
39061	gattatataa	attatagctt	caacaaacaa	tctcaatagg	ataattttcc	taaaatttcc
39121	atgaaacatc	tccattattcc	aaagagtaag	agggagagga	agacagaagg	gttacctctt
39181	tctgagttag	agagttagat	agccattatg	tgagcagggt	gtatcgctga	agtcctccaa
39241	taccagatgc	atgactctg	ctctggtcta	aggtgaaaa	ccagctccat	ggcagtgctc
39301	gtttccggag	aattggaaa	gagagtctgc	atgccagaag	aaacattgtc	atgacataga
39361	agtcgccaaa	aacataaatt	actgagctga	tccagtgtgc	atgtctctatt	ccaaatcttc
39421	caaaaacaaa	tagaaaaaaa	aaaccttctc	taataacaa	gggaattctt	tctactctta
39481	tgttctcatg	cccacacata	tacatggaaa	aagcattctt	gaaaaatata	cgatttgtctt
39541	agttaaggct	gctataacaa	agtaccacag	gtaagggtgt	tataaanaac	agaaatttat
39601	tttcgcgaat	ctagagctg	gtgtgtctat	atagcgtgtc	caacacagtc	gggttctgat
39661	gaaagctgtc	tataggctg	aagactccca	acttcttgtt	gtatctctac	atgcagagaga
39721	aaggggcaag	aacatattct	gggtcccttt	tataaggcca	ctaactccat	tcatgaggcg
39781	tctaaatggg	ggccatgac	ctaactcaet	ccaaaagact	ccaactctta	atataacaca
39841	ctgaggaatg	gtatttcaac	tgtgaatttt	gtgaagacac	aaacattctg	tctgtaaaaa
39901	taattcttga	tccggaatag	cagatgcaca	gtgaaattcc	caagtgttctt	acttaataatt
39961	aatttaagtt	tactatgtca	aataatgttt	tatcttagtt	tctaaggacca	ctaggacaaa
40021	aaacatgagc	gtggcccaaa	ggagttcagt	tctcgaggta	cccttttaatt	gtgtgttaaa
40081	ctgacaaact	ataaactgta	cataaagtat	aatttttctt	ttttttttgag	atggagttctc
40141	gctctgtcac	ccaggctaga	gtgcagtggc	gcgactctag	cttactacaa	ctgcgcctct
40201	ccgggttctac	gccatttctcc	tgtctcagcc	tcccaagtag	ctgggactac	agggcgctcg
40261	caccacacat	ggctaatttt	tgttactttt	agttagagac	aggtttctacc	atgttagcca
40321	ggatgtgtct	gatctctctga	cctcgtgatc	cgccagcatc	agcctcccaa	agtgctggga
40381	ttaacagcgt	gagccaccgt	gcccgcccca	taaaagtata	ttttaagttg	ttgataatga
40441	acacctgtga	aaccattacc	acaatcaaaa	tagcaaatgt	atgcatcacc	ccaaagtttcc
40501	tgtgtccctc	tctaatctct	tctctcacaca	tctctctctc	accatccccc	actctccatt
40561	gcttggaaag	ttttttgaat	ataaaaatgc	taagctaaaa	ggtaacacag	tcgaagagtt
40621	ctctcaaaag	taaggaaatg	ggaaggaaaa	ggtagcagag	tagtttcaat	tgtataagaa
40681	taagaagaag	tgactcaaat	atataatgat	aaatgatcaa	gtcaacagtt	tcaaagagcc
40741	aattggccat	acgactcaaa	aaaaacaaaa	gtgccacact	tatttaggaa	gggcacagga
40801	aacagaagca	cttctcttta	cacagaccac	tcaactctgt	agccatgtgc	aaattctgccc
40861	atcatcttct	aagtaataata	acagtgggaa	ctaaaggcag	tccaagaaat	aaaatgataa
40921	ttaaggacaa	ctaagaccct	tctcgtatag	aaatcaaggc	ccaagatgct	gccagctagg
40981	tctgttttat	cactggtgat	gctgacatct	ctttatccaa	gtctgtcttc	ctcagatgta
41041	gaaatagcca	tcactactaa	tttgttgagc	tttaatgaga	attaaccata	tggaaaccaa
41101	acacaggggt	tctaactggg	gggcttgaga	gtctgagttt	tgaagggcc	ctcgaatttt
41161	ttatatttat	atttgagttc	atagacttga	gcagtcttat	gtctccaggag	cttaagaacct
41221	actggcctag	gacatagcag	gcgttcaaca	aatataaaYc	attttctcac	ccatctctcc
41281	tcttttctct	caaatgttgt	cttgctgtaa	gtctatttag	aagaaacaYa	ctgtttgtca
41341	atataaaatt	gttttaaatg	atctagttaa	aaaagagaag	agcaaaagac	tctctatgat
41401	gccattttac	taagtctcct	atgtaagagt	ctgaatagga	agctgtcttt	tgaatttcagt
41461	gcataaaaaa	acttggccct	cctcgaggaa	ctggaatgct	cctctctctc	taaggaagaa
41521	atgaaattta	acatgcttta	ccttctgctt	tagctgcmaa	gaactttggc	aaattctgcag
41581	tatgtgtgac	atcggtatct	cttgttaata	atactattta	tctcttctgt	tttttccata
41641	gctgaattct	tccattttcta	gtctattaac	taccoccccta	cacacatacc	ccacaacaca
41701	cacgcacaca	cgtgcacaac	cacatgcaca	caattttccaa	acctctatgg	tgggggaaag
41761	aaggcaggaa	taaaaaatac	caaaaatgta	agaaaagatg	tgtgaccatt	gcccttaaat
41821	tatatatagg	gctggggcgc	agtggtcctat	gcctgcaatc	ccagcacttt	gggagtccca
41881	gacagcgaga	tcacttgagg	ccaggagttc	aagaccagcc	tggggcaacg	agcaaaacct
41941	catctctctca	aaaaaaaatt	gttaaaaaatt	agccaaagtgt	ggtagacatt	ccgtgttagtt
42001	ccagatcac	ttaagaccag	gaatttgagc	catgatcaca	cgaccacatt	ccggcgtggg
42061	tgtggagagt	agatcttatc	tctaaaaaaa	taataaaaaa	tttttaaaat	tacatatatg
42121	atagatttaa	gacagaattg	tttaccaaat	tccagaatat	tataatatag	gcttaggaca
42181	ataaaaaaga	agtataactt	catacaaggt	gaaatcgatg	tatttgctgt	ggcgatcata
42241	gagactaaat	gcagatagat	tgtaaaaaat	tataataagt	gcagactctt	catagcttaa
42301	actaaccagg	agaatttttaa	aaactttcta	agggctgagg	tggagtttaag	gggttttagtt
42361	atgacatcta	taatatagatc	cttaatactg	ctggaggcat	agcacagggc	tgtatgaact
42421	gcagcactata	atttttttgt	tgttttgctg	acagagtcca	aagaataaag	ttgatataat
42481	aggttctctg	gtaaagtcca	atactatagc	taaaacacaa	tgaaaaccca	ttaaacatat
42541	attctcaggt	aaaagtattt	aaggcccccac	taaagttaaga	aaatgaaagg	tggggaattc

42601 acgaagaaaaa cttgcctgca tagtgggtgac atttgcacaa ggggtgccctt cccccagct
 42661 cagctctccca ttaccccttt ttaggctcaga gactgcacaa ctcacgaagag gcttagagata
 42721 cctccagtgca ctaactctgaa aacatctcat cttgattttt taaaattttaa gacttaagat
 42781 cataaatacga aaagagatag agatactcaaa ttctaatgtg cttacacaaat accagagcccc
 42841 caaaataatag aatgattttt tactaaaaaa attcttttgc aacgtccagag aaaggaagaaa
 42901 agttgtatttc tgactctgtg agtgtatttg gaaaaaaga aaggtccatc tgcttgaat
 42961 attcatgcгаа атagtcttcc ctggaaaggg cctatlgagc ctgaagatc taaatttccc
 43021 caccagaaaat tcaataagac ctacaacgag ataaactttt catttaacat ctgtactct
 43081 catatcaatt tgcagaaatt atattcataa atgaaattgg aaattaaacc ctgagaaaac
 43141 tctcttgaggt aatataggtt attgtccata ttctactgct aaagacaatg tggcttaaaa
 43201 agtgcagcag cccatccagg gtcatccagg cagtgaaagg tagagctcag gaactccacg
 43261 caagaggtat ttcaaaagggt atgctgaaga caggctcctg gaaccagcag aactctccc
 43321 tctggtagca gcaataacaa tgatctctgg aacagagtg ctgaatttaa agatgacat
 43381 cctgtggggc tcagctcaat catgaatgtg cctgggttaa agaaccataa cctgtactg
 43441 aatcttgag gaactcttgg cogaagagac acagagtggt cgaatttaat accoggtgtt
 43501 accagctgtg tgaatttaag caggtctctt aatccctt aagcctcaat tctctcaat
 43561 gtctaatgga gactcctaa gaaagtgaaa tcaatacga caaaaatgtc caatgagc
 43621 ccacacagg cactgcacat aaaaattgta accacaggca tctctgaaa gagagctttt
 43681 gtttgcttaa ggaataaact gactgcataa aaagatagg ttatacaag aactgatttt
 43741 caaacacctc cccaggctgt tccatctctg ctacttgctg cacacagcgc catctttttt
 43801 actctgtccc cactctctgt tgatgtgttt ctggcagctg cctcagctgt ggtcttgatt
 43861 caactcaact ccaaaataat cctttaatta atgtgctttt actcctgtgc ggtcttgatt
 43921 tctcaaccca cccccatcc tacagctcag atctgtctgc atgacacga gaaaagattc
 43981 caagtctcac ataacagctc ttgggggttc aattctaaac aaacagacaa aagatttga
 44041 ttaataatag atgaaaagga attgatgaaa aaagattata attcaccga tctccatagt
 44101 ggaactgtca tgattttatt tagaaggtaa ttagagtcag tggcggttaa aacatttcag
 44161 atactataata acttaataac acattcttat ataccatgtc ccaaaaaggt caattttcaa
 44221 cactttttt ttaatacga aagctccact ggaatttta aaaccaaaca gaccaaanaa
 44281 aatcaagaat gagtagcctg tggaaagaca gccaaaaaaa agacacaaat agtgttaaga
 44341 ctgataataa cttaacctct atatttttc cctaaagctc tgaatgtac attacattca
 44401 taacttaagt tttttttaa aaaaatgatg aataaaaact tacaatttca tttctaaac
 44461 tctagagaaa attgtaatag ttttaacaar gacaaatgaa tagatcataa tctcttata
 44521 tggtagcaga aagacatttt accattttaa tggaccaaac tccagctgga acagagatta
 44581 aaaaataaag gaatggcгаа tagaattctg tagaggaagt ctggaaaaag tttgttgca
 44641 cgtgtttttt ctttttaaaa agaaaacaga aaacatgttt tataatttcc aggccttccc
 44701 attatcatga ttattttatt agaaaagcaa attatcataa aactataag gatttcagta
 44761 aattacacta aaaaattttc attttatgac cttcaagtgtt ggtcccatcc atcatgactc
 44821 cttcaagctc taaaaattcc caccogtatt gccagtatca gttcgtcttc tagaataatt
 44881 atcagatttct tatcctttca ctgttccacag cattcaactc actttactat ctgctctttc
 44941 tagtctacat tctcagtttt tcttttttgt tctcattttc actttatggc aatatttata
 45001 aggtactcat ggactaggca ctctagtgtg tgggcacact gatcaaaact gacacaaaca
 45061 aattcataaa ggttatccaa ggctctctga attatcactt aaatttgccc ttgtatata
 45121 ttttagaaaa tttctataa atcctcatgt ctatctccca tgacatgagc tctccatcag
 45181 aacagtataa agaaccttta aaagcccttg aaagagactga agccatcaac tagagaagc
 45241 ttcagtgtaa acttttctta accctccctt aaacctatca ccagatgag gatgaatcac
 45301 aactcaaaaa tacttgaga aacaaaaact tggaaagaca gtacggaaac agcaaacat
 45361 ggaacagtaa atacattgtt aaacaaactga tgaattgaca aaataagtag gggggatta
 45421 aaagcaggta tctcttgagg tatattttaa ctgatttttt tttttttttg aggtgtatc
 45481 tcaactctgac acctagggtg agtgcaagtg cgcaatctta gctcatgaa atttccact
 45541 ccaggttcaa gcaattctcc tgccctcaacc tctgagtagc tctggatgac agggactgtc
 45601 caccacgctc ggctaatttt tgatttttta gttagagcgg ggtttcataa tgttggcag
 45661 cctagtctgc aactcctgac ctcaagcgat ccgcccactc cagcctccca aagtgtcgg
 45721 attacaagca tgagctaccc tgcccgctg atttttcttt tttttacact tgtctcttt
 45781 tttctttttt tttttttttt tttttttgag acagagtctg gctctgtcgc caggtcggag
 45841 tgcagtgctg caactctagc tcaactgaaat ctccactctc aggttcgaagc aattctcctg
 45901 cctcaacccc ctgagtagct gggattacag gcatgtgcca ccaagctcgg ctaatttttt
 45961 gtatttttag tagggactgg gtttcaactt gttggccagg ctggtctcaa tcttctgacc
 46021 tctgtctccg cccacctggg cctcccaaa tgttgggatt caaagcatga gccaccagc
 46081 caggccaact ctacattttt aatgtttata aataggaac acacactaga aaaaaaagg
 46141 gacaaaatgg gacattctct caacattctg tcaactgttt tcttctctt aaatggttaa
 46201 tatttttcta gaaaatatat aatccttata aaagtgcta aataaacaga aagtcaactc
 46261 cctccagcaag tacaatgtac agaaaatcacc actgctcaaa gttgggatat actcttttag
 46321 gcttaggcag ataaagcat gcatatttcc gcatagatgg aaaaagatat aaatgtgtt
 46381 gatattgccc atttaacagg tcttatatat ctttccatag caacacagat aatgtactgt

46441 ttaccattttt aaaagctaca gaggatttcca ttctgggtat gtaccataat taccatccca
46501 attctccattt atgaacatc atctataagc ccatataagc aattataaag
46561 aaataaaattt gtcttgattt tgtgggaagc tatccagaa agcagatgtg tccaggcaat
46621 aaatctattg agagattttt tgcctgagca gtgaacaga tgacccrttga ccttaaaatg
46681 ttaactgtga ctaaaaaagt ctgtgtctag tctaagatat ccagaaataa aaacaaatg
46741 accatgtgag aatacacaat gattaggacc aaaaattat ctgagrtctt ctgtctgaa
46801 tcttctctca agaaatatac tgaattttca ggaataata caactttcat ctgttatttc
46861 ctgagtttcc caaaaagctac agaaaaatga tgcctatca tcttcttctc cctaaatgcc
46921 tacagtcgca accaggcaga agccaggaata aaaaaggag ggaactatga tacattatgt
46981 aaacactggt cacttgacaa aatagttagc ggtattaaq aggaagtatg tgggggtgtg
47041 atttttaagt attttttttt ttctcattac agatttcttt atttctttaa ttctatagt
47101 gaccataagt tgtccttttc ataagaattt ataatacagt gattatacac ctgagctatc
47161 cattacaatt atctgggggg gattaaaaaa aaacagaaac actgctgctt gggaccctct
47221 ccagatgtca taaatcaaaa tccctggggc taaggtccag gtatcaatta aaaaacaaat
47281 attcattaaa tactatgcag agaaatcacg aaagtttaca caccactgaa ttaaccaga
47341 cctaaacagc attctagcat cttagtggtc cccctgtggc cctctccaat ctctgctatc
47401 ttcttcccca gaggaaaata gggctcctaag ttcaaacact aggaatttat ttggacgaat
47461 tcaatttgaa agaatttaca gcacaatttt tgtgtgtttg gcttctttaa aatcagactt
47521 atgttttgtgc aattcattcc agctgctata ggtagctata actcattacc ataatgttat
47581 acttcattgaa aatacatgcc acaattatct gttctactgt taataaaact ctgtattttt
47641 ggttttagagt tataataaac tctaaacatt ttctgtgcat gttcttttgt acacgcata
47701 acacattttt tctgggtgga acctgtgagt agaatatca ggaattatga tctctatc
47761 ttcagtagat actgcccagtc ttccaacaa gttttgcaa tgtataccgc atcaacttac
47821 tgtaggaagat tgttccaatt gccacagaat ctacacata cctgttatca ctcttctatc
47881 tgcctagatc atagtgggat ctcatgttga ttttaattcg cacttcccta attactatg
47941 ggtgtttaaga aaacttttgt ttgcttattt gctatttggg tctctctttt tggaaacagt
48001 ctgttctgagt cttttacaca tattaataat ggaactgtcc ttcttcaatt taggaagttt
48061 ttaatactac atagatatga gttattctgac agatagcttg cgaataattt cccactactt
48121 tgttgcctgt cttttcactc tgtttacttt agattaacag ttttctattt agcagcaaat
48181 tctcaggtttt tcttttatga ttagtccctt ttgtgtcatt ttaaagaattt ctgttgcaac
48241 atataccaag atcatgaaga tacattccca tgttaacttc cagaagtttg ttttctctt
48301 cacattttaga gctaaaaatc agctgtagtc tgatttttac acatggttga agtagaggt
48361 cacagtctat tgtttcatat gttctatcac tgcacagcaa ccttcaactaa aaataccacc
48421 ctgttagtacc ccatgttctc tcaactataa atgggagcta tgaatttgtg acacatgggc
48481 ctacactgtg gactgataga cattgaagac tgagggaagt gggagggggc ctgaagatga
48541 gaaatttact aatgagtaca atgtacataa ttccagtgat ggaataccta aaagccctga
48601 ctccactgca caatataacc tgaacttaca ctgtgtctcc ataagtttat acaaataaaa
48661 taattttaaa ataccaccat ctcccccacc tactgctttg cagtctccact ttccataaaa
48721 gtcaagtga catatgtgtt tgcctctaga gtctcaattg catttctact gcttatttgt
48781 ctagctatag accacaaacc atacacaaat attcaactc tcttaactcc tgcctcaaat
48841 tgttctctcc tctcaaat tcttctgact atttctgact ctttactatt ccataataag
48901 tttagaaatc agcctgtcaa tatccacaca cacacaaaaa aacctctggg atttttgtt
48961 gggatggccc tgaatatata catttaatta ggaagaaatg atacctttaa aatatcttaa
49021 tcttttagaa ttgttaactr atgaatttcc gatattcctc tatttcttta gtttctcatt
49081 ttatctaaag aatgtttgaga tgtattatgt tattttctct gttagaggtc ttacatatcc
49141 catgattatc atccagaat acgttaagt ttcttgaggct attttaaagt actttatctt
49201 ttcttttaaa tttatttctt aacgttttgt agagatgag tcttgcata ttgatcagac
49261 tggcttga cctctggccc caagagatcc tcccccactc agtcttcaa ttctagggga
49321 ttacagagagc aaccacaga cccgaccctg ctttttcttt taaattgtcta ttctagctg
49381 tctcttccacc aatatatctt ttataaaaag ttctccatgt aattctaatg acttctggg
49441 ttgtgaactc cctaacaatg tgtttgataa acaagcaag gatataatgt tattttctt
49501 ccagctcttag ttatctctaa cctaactatc gtgtgaaat actttcttga tttttctg
49561 gcaatctaac atccattgta caattttttt cagaagattg ctttaaccaa accagtgtg
49621 tgttgatctc tggctgtgca ggaattttgtc atttttgtat ttccaaactg ttaacacagt
49681 tagtcaacta actggcatat agtaaatgtt tgttaaatga tgcctgaact atgtgacctc
49741 agcacaagta tgtacttctc ctgggcttca gttccctcaa agtaaaattg aggttaactg
49801 gctgccaatc gttggctgct gaaagtatca acaaaatggg aaacaaactt taaactggcc
49861 aaagctacac aataaagcat catttattaa ttctcatag ctttctgtg ctctacacat
49921 tcatctctct attgaaagaa aagaagaaaa gaagacaaat caggcaactc gagcagctag
49981 acaaaaaaat tagtgggaag aaaggaacta acatgtacag ggcatttcaa qaatttccg
50041 gcaactcata agcatctagc tgagaaagag acttccgaag agtcaacaa aagaatagct
50101 aaagtcagga caacttaaac agctcagctc ggtgaagact ctgttaattt ctgctcata
50161 aaacagattt agcaaaagaa gacacatac caaaaaatat cctgaagaa ctaacagaa
50221 ttaacttcca cctcaaggc atatggaaaa ggatttatta gaagcatttt atgaggcata

50281 attatattctt atgactttgca gaggtagggg aaaaaagaca gcacatcaaa agatattaca
50341 aagcaaaagt tcaacatcaa ttaacttagc ttctctctaa aagaagaagg aagaagaagg
50401 ctcttttttgc catctgacct atgtggggca ctgtgaacaa tttttccaac catctttcca
50461 tctttaccga gatacaattt ttctttccaa taattccact ttaaaaagag tgcgcccaatc
50521 ctaatactttt ccttttttct tccgaatcag atcctaactc ggcacacata aaaaatttcca
50581 actctctctt ttactatttt tactaaaaat ttcttctact gttgtgtgtt aacacacaaa
50641 accaaatctt aagagaaaatc catcatagcc tagtcttcat aatttctctt acattaaaaa
50701 caaaataggt tcaagtctact gaggttggca gcagtcaacc gccctcccca ttcttctatt
50761 gttatctctt tcaaaatggt caatagccct aaagggtgaa aggggaacata aaaaatgaaca
50821 ttatagctct acacactttt tagagaaaag cagactttgc aactaaccat gaacagccgtg
50881 ctactcttgt aaagcatcat caatagcaag aagactgaaa ggcccaaatg atgctgtgct
50941 caagtccaat tactatgaaa ttccacggga gttactggat atgaattcaa acaaatgaaa
51001 gatgttaag gcagcagtg actgggactc caagsgaagg gaaactgctc taaaagtatg
51061 atccatctat tgtgcttact ggtcacagaa agcaccctcc tcagacattc taggagactc
51121 atctgccacg ttccaagatt taattagcat tcagggccact acagaagtca tgcgaattag
51181 ctacatctgc caaagatggg taaactctcc taattacag cagatgtcca atttctcag
51241 ttgttatggt tatcagaatt atcatcatca aacatcaact cagcctattt cttagggagt
51301 tatgcaagtc ctgaactatg tcttctaccc ttctctactc gattttttgc tagatagatt
51361 ggagaatcat agcttttatt atcaactcagc agttaagtcc ccacattgtc cctctctctc
51421 acagatcagac aataccaatg caaggcaaaa agataaacat tgaggaaatt tagactcata
51481 atcatagcaa agaggcaaaa agcagtgctc gaaaaaaatt aatttctcaa ttttctcata
51541 ttttttcaat ttgtataaa aatttctagg tatataact taacaccaag taacagaata
51601 aagttcttga cgctgctctc cctgactcag gggtggcctc cgatacaacta aaaaactcaa
51661 taactccaat tcttcacctg ttaaaaatga gatagcatct ttctctactc catcataagg
51721 ctatttttga gaatgaggtat tcacagaagg attaaagtaca aagaccatgt catgtgaaga
51781 tattcttatt ggagacagac tttacggaa ggaaggactc atgaactttg caggagactc
51841 cactctgggtt caagtcctat ctgtaacctc agcagctgtt catttgttct tctgacagg
51901 gaggaaagcat attacattgt aggggtgttg ggatttaaat acactgcagc tggaaagcag
51961 ctggcacaac ctatacagag tgaatgttag tgcctatta taagcaaggt ttttctgcaa
52021 aaaaaaaagt tgccttttcc agctttggcc tcagctctat acataaagg cctagaattc
52081 agtttgcaca agggcttttc aaagataagt gcatagcaac acaatcaact tttttagaaa
52141 atattatggt tcagaaactgc aactagtcta ttaacctcaa aacaaaaagc aatgtgcata
52201 gtttcatgga aaagcacatg atgactctgc tggcacaact agtgcctttt acattttata
52261 ttaacaaaacc atgagctgtt gcccaaaccc attccaag cacaagctt acaatctaac
52321 atcccatctc gaggttctcc tgagatgaac aagaattacc tgaattcgcc cgggcccgtg
52381 ggctcatctc tgaatccca gcactctggg agggccaagg gggcagatgc cgaggtcagg
52441 agttcgagac gaccctgaac aacatggtga aaccccgctc gtactaaaaa tacaataact
52501 agctgggtgt ggtggcgggc gcctgtaact ccagctactc tggaggctga ggcaggagaa
52561 ttgcttgaac ccgggtgtgc gaggttgacg tcagccaaga tcatgccact gcactccagc
52621 ctaggcgaca gaggcagact ccgtcaaaaa aaaaaaaagc aaaaactcacc tgaactctc
52681 aaaaatgggt tactcatagg gagccaagcc cggtcatgtt gcttctgttc tcccaagggt
52741 gaatttgaag tggcaataga actctgttga ttctgttttt tctcaattta ttcttactct
52801 ctattaaatt gttatttttt acagttcatt tatggtagac aagttatttt ctgttggtag
52861 ctcaatgat tgggaatttt gtgatctctg taaacgaatt aatctctcat agagatctat
52921 ttctctgttg atatgagaca acagagaaac taaggagaac aactctgata cttttaggt
52981 caaattcgatt agaaatttct tgcctcaag atttgtggct gtaatttcat atttctcatt
53041 caagcttttg tgtttctgtg ttatttttc atttgtggct gaacttgtag atttgaggt
53101 atacacattc tatctgtatg tctatgtatc tgaattcag aagctcctt atttctcatt
53161 gtgagtatg aatgttttcc tgtctccaga tggattaat aaattagatg gttaaagctc
53221 tttaaaaaa aagcctcatt atgggttcat ttctcctctc aactctgta aacatcaacc
53281 agtttgaaaa gYctctagat ttgggggtgt aaaaattcca tcgagcttga gagccatcag
53341 tcttaagcgc gctagttcca ccccaagctc tgcgtccagc atgctcagac agYccactgg
53401 gcagcaaaag ggcgccttag taagtgaatt taaagccaca aacatggctc tctccataag
53461 gagaataac tgtgttaaaa atttatctct atttatctct ccgctgcaag tccgctaaag
53521 gaggagggtta atagtaagac ctctctagcc atgagtggtt ccgctgcaag aaaaactcacc
53581 cgtggcggtt ctgagctcac caccacagct gtaccaacgc cccagacacc atctcctctc
53641 tagcgtgaca gaccctggc ggaatttgac ctgcgaagac atgcaaaaagc aactttgggt
53701 tgcagaagc accttaccct cgtggaatt cccgaagac ccgagctgc tccgctgac
53761 cagccacccg gtgttgggat gccagcaccg ctacgcaccg cagctgcagg atccccattg
53821 ctgctgggic cggcgtgggc ccaagcagc gctggggaga ggtgcggcgg tggcggcg
53881 gggcccgagc cctcgccccc gctatcccg ggcgcccgca tcggcgccca tcccgccga
53941 ggtgcagagc ctctgcacct gccagctccg cagcgcccg cgcgcccgcc gcgcaatcc
54001 tagccccagc gagaggtttc ctcttccggc acctgacccc caagcgcgtg tcccgccgt
54061 cctccgggc tctcgctgcg tcacagcggc gggtctctgg cgctggggcg cgctggggcg

54121	tttccaggca	gcgcgcgcct	cctgccggcg	cgcggggaac	gcgcagccgt	cgggtctccc
54181	ggagagcccc	aggaagtcct	ggcttgagcg	ccggtctctg	cccactttta	tctcaccaga
54241	taggcaaaac	cactgtgaaa	gtacaaaata	actttctgtt	aggagaaata	gtccgccagcg
54301	tccgttttcc	gtgtttaaac	ccgctttaca	acagcaaaaa	agaggaatac	gttgcgtaaa
54361	acacaaattga	aggttatttc	ccccgcgtag	aattgtcttc	gctgtcgtgag	cttatgttag
54421	caataaaatg	ccaggagaaat	gctgtttgtg	gaattgttcc	cgactagctg	ataagcacta
54481	catttaaaag	tagtttcgtt	ttctctaaag	gaattatctc	cgaaagcacg	gccttggctc
54541	gaactctctg	tcaagtctca	tacctgagtg	ctgtttgagg	cgactgagtg	cattgtgacag
54601	atgctgctgc	gatgatgtgg	aattagacgc	atagacttaa	cgttatttgg	ggagagtggt
54661	ctgtcttctc	ttccctgtgt	ttaatttttt	tcaactatat	attttgtgat	gcactctata
54721	gcttggaaat	tacatgttca	gttagtccac	tactaagcta	acaaatata	gctgtgacca
54781	taagtgtgaag	aagtgttaaa	gacaaaatta	ttcaatgata	cttgtttaaag	caccgaancg
54841	agactattga	ggactgtctc	aattggcata	gggaccactg	caactgggtc	ttgcagtggg
54901	gaagaggagc	tgggctcaac	tccgaataca	gcgtgggcaa	gagggaaatt	atagccaaac
54961	agcagtatgg	agttcagttg	atggaaaaat	actaagagga	agcatcacag	gtaaggggca
55021	tttggtgaaa	ctgacctaac	aggaactctg	aagacataac	accactgggg	gagtgttgca
55081	gggtgagaaa	actgattaga	tatggaggaa	agggggtctt	ttgtcaaaa	tggattttta
55141	caagaagatg	ccagagtggc	cctagcagaa	gattcagaag	cgctactaga	gtttgtccca
55201	gcataaattc	tttgtcatag	gcacaaatct	ttctctgctt	tcccactgct	tggtccaatt
55261	ttttcacact	ttgttagcac	tattctattc	tgtagtgtct	atttactctc	cacagtcgcc
55321	ctatttctat	aacctctagt	agatgaaatc	caaggaaagt	gaaaggggca	atttatgata
55381	gggtgtcatc	tgaatcagtg	gatggattat	tcaataaatt	tgtgttttaa	ctgtattagt
55441	gcactatrgg	cagagagaa	cagaaactca	tctcacata	attgtgaaaa	caaaaaccca
55501	aaatgaataa	accataaaa	agctagcaga	agaggggcgt	ttcttttttt	tttaataa
55561	tgtatctaaa	atgggaggag	ctttctaaag	agcaaaaacg	caaaaggaaa	aaccataaag
55621	gaagaagatg	gtagacttgg	ctacgttaaa	gctttttaca	taaatattta	attttttata
55681	aagttaaac	acaaaacata	tgaaaaagca	agttgacca	aaaaaaaa	tatgcaaa
55741	acatgaatga	ggaatttcc	cccaaaaaa	aaatacaaat	caaacgttaa	ttacagaaaa
55801	aagttcaact	tcaactagta	acaaaataaa	aaatacaaat	aaacaaacta	gatgttaaca
55861	tttttaaaag	ataaaacata	cagtggtgat	aaaagagcta	aggaaggagg	atttcttagt
55921	agggataact	ggaatataca	ctttaataaa	gtaactcgga	tcaaaaaatg	tttaaaactt
55981	tatttttggta	gattttggat	aatccacttt	tagattttaag	tacataata	ataaacacat
56041	taaaaaattag	tataagctgg	gggaagtggt	atgcgcctgt	agttccagat	actcaggggg
56101	ctgcgtcagg	aggattactt	gagccacaga	gttcaagagg	cgactgtctg	attgatggcc
56161	ctgtgaatag	ccactgcctc	ccagcctggg	caacatagtg	agaccccatc	tcttaaaaaa
56221	aaatagtgtg	cagaactatt	catagagtat	tatttagagt	agcagaaaag	tctaaaaaca
56281	cctaaatgcc	caacaattag	gagttaaata	acctatgata	gagctttata	agaggtactc
56341	gtgcagccct	tcatttatat	tctgattgaa	gtctttatga	aaagccaagt	tatgatgcaa
56401	taagtatatat	tagtacrtgg	cttttttttt	ttttNaacat	acatctttta	ttgactcaga
56461	ggaatggata	cttttgggtg	tgtttgtgtt	ttttaaatgg	agttcttttc	tgtccccag
56521	gctagagtgc	agtggtgtcaa	tcttggctca	ctgcacacct	tgcttccccg	gttcaaggca
56581	ttctctgcc	tcgcctcccc	aagttagctg	gattacagcg	acacacaccc	attgccggcg
56641	aatttttggta	ttttttttta	gtagagatgg	ggtttccaca	cgttggccag	gctgggttgg
56701	aaactctgac	cttgtgatct	gccacacctg	gcctcctgaa	gtgctgggat	tacaggctgt
56761	agccacacca	ccggcgctgt	aatttttttt	taagagacaa	gatctcacata	tgttaaccag
56821	gctgcctcta	aaactccaga	tgtaaagaat	aaagaagaaa	gaacacagaa	aggttgctta
56881	gaagtcaagg	acaagtttat	tttagagaaa	acaaaacctga	aaaggggctt	tggccagtat
56941	ggttagagcg	acactctctt	acggactaat	aaaggttttt	aaagatttag	ggtggggagg
57001	tttatcagag	gcttgagact	cttctgtgtc	tgttgtgctt	atctggggag	gagagttgtg
57061	tgtctgttcc	catacatctt	tttgcagctg	caggcatacc	cccgactctg	cttttagctt
57121	ccctatctta	gtgcacctga	agggaaaagg	atgtacttac	taaggccac	ttgtttactg
57181	gggccatctg	tatgagggtg	aagtttggca	gttaccacaag	aagatttccc	ccacactccc
57241	tttgtgcctg	agctgtctta	tctgtgtttt	actgtctgtc	cttctgtcgt	gcttatagtt
57301	agaagagagc	tgatttctct	gaaatgcagt	agggctagaaa	ggagagctga	acttaaaagt
57361	cggtgttttg	tccgagatga	cggtgtctct	ggctgtctac	cagactcaag	caactctgcc
57421	acctcagcct	ccagactaac	tgggactact	acaggtctac	acactgtgct	gcttctgtgt
57481	ttcttttgat	tatgcattcc	tggttatgta	catagaaaaa	aaagtttgaa	aatatatgct
57541	aaaaatgtta	aaatagttat	ctttaggtgg	ttaaattaa	aagattttcc	ttgtcacatt
57601	ctctgtgttaa	agttctacat	tgaagtgtat	cattttttaa	taagaattta	tgtatttcta
57661	aaagatacat	ataattttta	aggttttctt	ataaaagcat	tcaagataata	agtcacactt
57721	cctgagtatg	acacataaat	gcttttttgg	atgcaacctc	tgtctacact	cctaggttgt
57781	tgtcgcaact	aaaaacagac	ctctacatgc	tggccacatt	gaattatttt	caatttctct
57841	aactgagcct	attctctctc	ctccacacct	tgtcatcgca	gttccccagg	cctcaggttg
57901	cctgtgcctg	catgacagtt	gatgacaaat	gtgtcagcaa	tcagaaagag	acctctatc

57961 cttggcttct aagtacagta ttgtgacagc ttgggaggat gattggctgg gttccatgac
58021 tcatgtcaga ttccattgg agagctgaga agcaacaag cttagcagct ttgcagatta
58081 gttcattttt cctatgtatt gaagagaaat gcagaaactc aaagattccc aactccaaga
58141 tatcatcat tagagcgtta aagcttctgt aggcaggcgg cagtggtcta taactgttaat
58201 cccagctatt tagcagacca aggcaggagg atctctggag accagagagt caagatcagc
58261 ccaggtacca ttgtgagatc cccattctta cacacacaca cccacacaca
58321 accagaaaaa gaaattagct gggcatgggt gggcacatct gtagtccctag atcacaggt
58381 gctgaggctt ggaagattgt tcgagcctag gaagttagt ttgcagcttag actcattat
58441 gccactatcc tccagcctgg gtgacagagc aagaccctgt atctaataaa accaaaacct
58501 tgttaaaatt agaagaggaa ggaagggcag gaggagaga gaggcgaagc aaaaaagctt
58561 acccttagctt tttaataaaa acctaaaata ccttaaatgg agctgggagt ggggactcag
58621 ggaaaaggca ggtattgaacc tctgtgaaat ttgccattc attctttccc actctaataa
58681 caccctctat ctgtattttt taattcttat gctcttttgt gggtagtaat ttacattctt
58741 tgttctatga gtaaatcca gtaaatgctt gtttagtagc tattatatgtt aaattattat
58801 gctagaaaaa tcattcattt ggttgataaa tatgtgccag gctaaaWgtt aggtgctagg
58861 tactgaaaaa ttggaacacg aatatgtttt ctgctgcaaa ctgacccatg acttgctggg
58921 tgaattagct tgaatgacaa ataattacag cacagtatgc taagtgcatt acagaaagt
58981 taagaaacca ttgtgttaaa gattgggagc atttaacttt cagaattcag caaaggcttt
59041 tataaagaga tgtttaaatt aagattttgt tcacttagtat agtggcnaat aaaaaaata
59101 ataatagacc ggcacagctg gctgacgcct gtaattctag ccttttggga gggcgaaggc
59161 ggtggatcac ctgaggtcag gaattcaaga ccagcctggc caacatagca aacccccctg
59221 tctacacaaa tacaaaaaac ttaccagggt gtggtggcgg gcacctgaa tcccagctac
59281 ctggaggctt gaggtaggag aatcactaga accctggggg cagaggttgc agtgaagctga
59341 gctgtgcga ctgcattcca gcttgggtta cagagtgaga cagagtgcga ataaataaa
59401 taaataataa aataaagttt gaattactgt tggacaatca ccaaggggagc aaagtggaga
59461 ttgtggatga agttaggca tttcatacaa ggaagaaata attaggtgtt ttttatatg
59521 ttgtgcaatt ttgtaaaaa cagagaacatg tttagggaaat ttcaaagagt gtgttataga
59581 aaaaatgctta tttayagtta tctatcccgg cactaagaac agttttagata ttaataaat
59641 tgcaactcgt agtattaaag Ktattttaaaa attgctaagg gtgtttttct taaaaaatg
59701 ttgtctacgt cataaaaaac aaagcattta caaagttgac ttgagtgtct actctgaag
59761 ctgtaaaagg tatgcgaggg ttactataac ttgtctacttt tcttatattt tgaaactttt
59821 aataataata agtttcaaaag aaagtactca ccaaaattgt gtttttataa tatgagact
59881 gtaatactgg gttcctgctc ctaaaaccat ataattccag ttggcagatga acagaagcaa
59941 attgaaaagc atatgagaaa gtgtccaaat catgaaaagg agaaaagctg ggtccctatg
60001 gactctagca attcataaaa ttatatagag aagtgtacag cttaaaaagg tagggtaagg
60061 attaaatagc aggaagaaaa ttgctttggg aaagggtaaa aatgaggaaa gagctagacc
60121 agtggttctc aattgggggc aattttgccc cccagcaaca tttgggacata tctggagata
60181 ctttatgttg gcataaatgg attgggggga ttgtaccggc atctcagggg taagggacag
60241 ggtatgctta aacatctctac agtgacacag tcagcatccc acagcaacag ataattctagc
60301 ctgaaaagtc aagagtgcag ccaattgagaa atcctgagct agatgcagag actayatttg
60361 cgaggtgggt tattcttgaa gtctacagct tcaagagctt cctaaaMccc ttttcaaatg
60421 taggtatagt tgagagtctg gctctgggaa aggtgtccaca gcttttatct gtttctaaa
60481 aagatccgtg gccgaagag gctgaagaac caggaaaaacc aaggggtcag gggactttgg
60541 tgaaaacagt tatttcaact agaaaaattat caggaggtga ggcgtgtcag ttgttctgaa
60601 gccagattat agtaaatctt gaaaaaccag cagagatagt ttgacatcaa cctgaggtct
60661 ctgtgagagc acagaagacc tattaggaa ataatggcag gacaaaagct gttggggcga
60721 atgaattcgt tgaagtggct tctctacttc aactgcccaa gtctccctca atcagctctg
60781 aaaaactagt attcttatgg agtggaacta actatatcca actcatccc aaactctatt
60841 aatttttttt ctaaaaaata tatttttagct tttctttccc cgtctatcag ggtcaacact
60901 ctattctagg cccagctcat cttaatcttg gacaaactga aaaccttctt tagtgactc
60961 ttggcttat catcttttga aattgagcac ttgatatact tctgtgagc tggcagcat
61021 tccgtaagct acactctcaa gtgaggtgg gaggagaagg cttatgggga tcccgcaaat
61081 acagtgaaaa caaattacag agtgatttct cattagctcc attatgggga acatgctca
61141 tccagcttct taatgagctt ccagtaagtc ttgagactgt ttggggctga ttgggggga
61201 ttggtgtgaa gcagaagtgg tgcttcacag taagtgttgg gataaatcca acagactgag
61261 tctcccaggc actgtgtctt ccaccttccc cgcagccaca ggaagtgaa cccagataat
61321 cagcctgaga aagaggatag gagagttttt tgaagctcac aaactctgt ta ttgctgctg
61381 ccaaaactat tcttaccag actatcctca ggctaattt ccttaaatca acgtttctga
61441 ttatcacttt tcacacttaa actctcaata agtcccaact ctaaatctaa atgtgcagc
61501 taatttttct tggccgagca cggttggctc cgcctgtaat ccagactctt ttgagggcga
61561 agggggggcg agcagaggt cgggagatcg agaccattct ggcataactc gtgacacccc
61621 gttctactga aagatcaaaa aaaaaaaata aagtgcggg cggtgttcgg acccgtgag
61681 tccagctcac tccggaggct gaggcaggag aatggcgtag accgggagg cggagctggc
61741 agtgagcgca gatcgacca ctgcactcca gcttggcgga gcaaacgaga cctcgtctca

61801	aataaataaa	taataaataa	aataaataaa	taaaaaataa	tatatattact	cttcaatttg
61861	gtcaataata	aaacacgaga	gggtacgaga	aattacttaa	aaaaaataaa	tccccgcaga
61921	atccaaacat	ctctgacgga	tttaaacact	aaacaaccca	agggggagat	tccggctctt
61981	ggactgtatt	tcaaagcgat	tggagccag	accattccc	caattactca	aaagatccaa
62041	ctctctctca	gggtgtaccc	ggtctagaaa	ggccgcctcg	ggagatccac	ccactctcta
62101	cggggggggg	agttctctcc	gaacccaccc	gttccaagga	cgagagccctg	tgtttggcca
62161	gttagcgtcg	cagagctgcc	gactcgcgga	agagcagctg	ttttggcgctc	tgtttgcctc
62221	gttaggaaga	tggtggtatg	gctcagcgag	cttcaggatg	ctgtgaattc	ggtgaggaat
62281	ttaattcgat	tagcctctgt	ctttctcttt	cttgaggtaa	agcaaggagga	caataaggcc
62341	caagaggcaa	aacttgraa	gtgggagcgg	actgagagac	agggcttcgg	gaaagagccg
62401	tctctcgcca	ggcgccaccg	cgaactcgga	ggtttgagcc	ggcgctcgccg	gcggtgcttg
62461	aaggtgcggg	agggagcaga	ctcttctcgt	tgaggagacc	gggaagatcg	tcttcttcgc
62521	ggtagcgtcg	agggagcttg	tggtgctgcg	ggctctgctg	tgcttccctga	ggacagtctt
62581	attttcttgt	caagtgtgct	tacgggttct	ctttcttttg	gggtgttgga	ggagagcag
62641	gagctttttt	cttctctcat	caattttaa	ccaccaagg	agatctcatg	ctctgagatg
62701	ctgggttttt	gtaatgaaa	gtcagccgag	agtattctct	ctccagcgct	cttcaagtgtg
62761	tgccggagcg	atgctgaggt	gaggtgagag	attttcttct	atcttagctc	gttctctctg
62821	gccacagata	tcttccatc	tcttgcctc	tgccatccc	tcaaaaatac	tcagttttgg
62881	gacatcttta	tgatagaaa	tcttctaact	ctcgtatgc	tttccgctta	tttgaagctt
62941	ttttcttttt	ttttttttt	tttttttact	ttggagacag	tgcttcttgc	ccagcttgga
63001	gtgcagtgcc	gcgatctctg	ctcaactgca	ctctccgctc	ccgggttcaa	gcgatctctc
63061	tgctctagcc	aaagtgtcgt	ggatttacagg	cgccctccac	cacgcocggg	tattgtttgt
63121	attttttagt	gagacgggg	ttcaccatg	tgctcaggct	ggctccaaac	tcccgccttc
63181	aaacgtatct	ccgcctctgg	ctcccatag	tgctgggatt	acaggtgtga	gccacggctc
63241	ccggcctaca	agtccttttt	aatgtctccc	ccaaattcta	ctctctatta	aaggccaagt
63301	ctagtataa	gtccccacg	tttgttctct	ctgtttattt	ttctcttcca	cttagctctg
63361	gttggaagat	gaagcaaa	cagaggatgc	cgactcatgg	gccccctaatg	aagggttgat
63421	gttctgagaa	atgcattggt	ttggcgataa	cttttttttt	tttttttaata	atgtagaatt
63481	caatgtaaaa	taacgataaa	acagtaatac	ataaacagct	aaacaaagta	tttagtatca
63541	tattcaagt	tcatggactg	cacataattg	tattgtctat	cttttctcat	cttagctcga
63601	tggttaggtg	atttacagca	gcataccacc	aagcactaag	gaacgcattat	cgctcacagg
63661	ttagtactgc	tacagtgta	ctaggcaata	agaatttttt	agctttatta	ttaacctaat
63721	ggactgacat	aatatgtgca	tcccgttgac	caagacgtca	gtgggttattc	ccatgogcca
63781	atgggtgtac	ttattaccat	tacgagttag	attgttttct	gtgaataagt	tttccattga
63841	tagttatgtt	tctcattgtc	tgattgtgat	ttaatgtctc	agataaagg	cttagagaaa
63901	aacatgctga	taggaaaacc	taagtacctg	gtttctctgc	cagttttttt	tttctcaact
63961	gtaagctaat	aecttggcat	tttctttaat	acatcagaat	ttgtacactt	gaagctgtgt
64021	aactctactg	tgataaaa	gatctacctc	gggtgaataa	cactagtttt	taagctgaga
64081	ggaccagcgt	ctaagtgtgt	aattcaccaa	gataaattct	accctgcagg	gacttgcctg
64141	ggctgcagcg	cagccttctc	cagggaaacta	ctcctttgag	gtttttacgag	atacaattgt
64201	aatgatgtcg	ctggatagaa	atcttatttt	agggaaaatt	tgtcatattc	atacatgtct
64261	tgtaactact	aggataataa	ggacagagtg	aatatggcct	taactccagg	ttaggcaata
64321	tgataaattg	attttggaga	aaagccttta	ggataactag	tatgagtgcc	tactgtgtgt
64381	aaagcaccat	atgtttacaa	agagataact	ctaaaggagg	ttaaaaaatg	cttaataact
64441	aggagacatt	taagtgtatg	aaaaaacagt	tggtgttagg	gagttataat	gaattgagta
64501	ttttcagctt	caggttctga	tgtaaaattg	aatatggagt	gatgtgaacc	aaattgagta
64561	gaatctctct	agcatttatc	ctaactcaga	ttatttgtat	aacttttaac	atttgttcta
64621	ccctctgaaa	accgtgagtg	cagaagctat	gtcgacttca	ctatgttctc	ttagtgttta
64681	tatgcagtat	tttagttatt	ttccataaat	tttgtttgac	taaatggatc	ttaaaaagct
64741	ttgaaagatg	gaaaagatag	acactgtgtt	gaattatttt	tttgtgtgtg	ttctctcttt
64801	acaacaatgc	tataatagga	gaatgtatcc	aatttaaaaa	tgaaaaaatg	tgctttctgt
64861	aataagagct	ggaatcagat	ctaggtttgt	atagactctg	agcaacttat	gtcccttatc
64921	aattagcaaa	acctggggca	agccatttag	tcataagtgta	ctttatgact	aggtagataa
64981	atattgaagga	gtagtggaaa	acgagtcaat	agagttgact	tgattatttg	aaagcttcaa
65041	aacagcctga	gttgagattt	tatcttgaag	acagtgttgg	agtaagctta	aatgttttga
65101	ctcaggttaa	ataaaatacc	tggtgaactg	gaaaaacaca	agaaagctta	tgtaataatt
65161	ctggtgtgag	ataattgtgt	gaactaagca	acagtgtatg	aagggaaaaa	gaagagtttg
65221	tagtatttag	agaaatgaa	gatcaaaagg	aagtaagaat	agtaagcttag	atttggttgc
65281	taggagactt	tcaatacaaa	ttaaatatta	aaaagaggaa	acagttgcct	ttgagttctg
65341	tactgtaaac	aatggcagag	aaacccagtg	gaataactta	ttgaaagtat	gttaggttag
65401	aggtcaagac	taggtgagag	aaatttgtat	ggaagtagtg	tgattttaat	ttgagagatg
65461	taaggaaatt	ctattcccaa	tctttaaaag	tacttaagga	atacttttag	agtgacaagg
65521	tttcaaaatt	gagctctgaa	caaatacatt	ctctcttact	taaggtgctt	tcttgagcac
65581	ccctctcttc	actatttctt	tacacttgta	actactccag	tctcagaaaa	ggaatcatat

65641 accttgcgtc tgagaaataa attttttatta aacatgggtgc catagatatt tacaagtcac
65701 attttgtctt ttccatcact atccccaaaa agtctttact ttgtggcttact aataagttact
65761 ggcccaattatt ctcttatcat tttactatca caagaaaaag aagattattt tgtatttcaa
65821 ctgtttccac tagattttata tcaaaattttaa cttttacctag attgatacca aaccttagtt
65881 ctttgcagaa ttaaaaagctt taatcaacat taataataac tggctttatt tatcaacaat
65941 cactcttttc ttgttttattc cctgtctcag agattatcat tccagaaact tatgttctt
66001 gccattacca ataaagtcct taaaaactgat aagtcctttc taacctttgat atgttttctt
66061 tggccatacc ttgcagatca gttttgtaat gccattggag ttttcagaca attgtgtctt
66121 cctgtcctct tcaataatat tcagacagaca attaacaagg accagccagc tgaacctaca
66181 gaaggaatac aggtttttctt agcttctctt agtttgactc tccatttttt tttttactct
66241 taaaattaga ttctttagaa aattcagctt attttatttc taaaattatt gtgttttagg
66301 aatatttaca gttttgtttg ttgtaaaaaa ttacagatca attcaactaa catctcataa
66361 ttctgtttca ttgcagataa tagattttga ctcttttaga ggagaatttg tatataatta
66421 gctgtagtag ggaccttaac aaggaaggca tccattttat aaattttctt attcaggtga
66481 ccaactatatt gtcatataaa atgaaattac atacattatg cttaattccc aatttgatg
66541 aaatgtgtaa caggattggg gtatttttta atcatttaga tcttcttttt ttttttaga
66601 cggagtctcg ctctgtcacc taggtgtgag tgcagtgat caatcttggc tcaactgaac
66661 ttctgctcct cgggttcaag gaattctcct gctcagcct cctgagtagc tgggataca
66721 ggctcatcgc accagctctg gctaattttt gatttttttg tagagatggg gtttccacct
66781 ttgtgtcagg ctgttctcaa actcctgacc tctgtatcgc cctgcttttg ctccccaag
66841 tctgtggatt accaggctga gccaccaagc ctggccgaat cattagaatc ttagggcaga
66901 aaaaagctata tgaagtttct ggtttctatta acttaaaatt tctttgttaa aatttgagg
66961 ttcttaatat cctaactag cagttattct ctctagatga tggccagctt ttgtgcagac
67021 tgaattgcag aacagcaaaa gacatttgat ttgtgata ttctctacc ttcttaccac
67081 ctacagctgc ttacaggtta agcctctatt ccttttagaa tttaaccag aatagaat
67141 ttgtaatba agaggtaga tttagtacct ttgtctgtg tccatttttg ctgctataac
67201 aacatacctg agacagtaatt ttataaagaa cagaaatttta ttcttccaca gttctggagt
67261 ctgggaagtc caagatcaa gcaatttatta ctcaactgtc atgagtaagt attttttatt
67321 aacatgggtt ccatgttttag ggccatgttg agggccttct tctgttttct tcaatgggt
67381 gaaggcagaa gggaacaaaga gcaagagctc cctctcaac ctcaaggcat attttaagg
67441 tactaatcc atctatgaga tctctgctt catgacttaa tatcctctta aaggccctt
67501 cctcttaatt accatcacat tggggattag attttaagt tgaattttg gaggggatc
67561 attcaaatca tagcaccttt ttgtataata ttcaaatct cattataagg aaacctctta
67621 ttctctcaa ttgtctctt ttctcagtg aagaatgaga agagggtggc atttcaaat
67681 tggagtga aaatgtgttc taaacgtaat gtgtaaggga gtcaactaaa gatactaga
67741 gatctctaat tgaaaagtc ttctttttaa taattcacat ggggaaatt cttaaaagt
67801 gtctgatttt tagaatttct ttactgaaa acatataca atgattgcaa agtcatccag
67861 atttaaaaa aagtaactct gtgacagctt ctctttctaa actctaaaga tttaaaatt
67921 tgtgtcttat aaaggctgct agcttgata agctagaaga agaaaacct gaagctgcta
68041 catgtctgga ggaattgttt tatcgaggag acatgcttct ggagaagata caaagcgac
68101 cagactcata gcatcagtg atcctgaaga caagaagtg taccatagc cagctctctc
68161 agaattctgc atcagactta gatcaagcc ttaccaaca ttacagaaac attaaacct
68221 atgacactK acccttttag ctatttttaa tagtcttcta ttctactct attaaacct
68281 ataaatcat gattgaatca gctttaaag atcataccat cattttttaa ctgagtga
68341 ttataagc atgtaataca ttaatgaaca taatataagg aaacatttctt ataatctgt
68401 tatgacataa ttattgtctc cattttgttg tattgtccag tacttttaca aatcaaaaca
68461 tctctcaag caaaggagaa gacagtaaga acagacataa gggacatttt aaataattat
68521 agtgtctgc ctcttagag tggcattgtg taaatctgag ctttagagaa gaagattttg
68581 aaaaattgtg ttgttttag aaatagagtt agtgtggctg atgtagaag acacattat
68641 gcaaatgttt ctctgtcag aacctcatac ctgttctagt tcaatccac gttttattta
68701 caaagatgc agtgttcaa agaatttctc ttcttcttca tcttcaattt tctacatct
68761 gaactatttt tctattttct ttcaactgt ctttttctca tcttcaattt tctgaagtt
68821 caaatgaagt atgttttttg aatcatcaat ctcttctca tcttcaattt cagaaactat
68881 cttttgctct tattaacctc tcaactttta ttgtcttat cttagtcagc aataattat
68941 tctcaactca tcaattactt ggcatcatt taattttgac taattttgac taagYtact
69001 ttgtgtcagc attttcaact atggttattc atccaacctt ttgtactctt tctattttcc
69061 gataactctc actggaaagg ttgaattgaa atgtgttcca catttttaact agactattcc
69121 tggggctgta tttttcagaa tatgtctgag tacaatttgc gcaattgtt gctattttga
69181 gtttaagttt caggagaaaa caggaaacaa tatgtttcaa atttcaggaa ccttcagct
69241 tcaagctttt ccaagtttct tgaaaagta atgtcatgaa aattttgtga aaactgtac
69301 agctgtaaa gtgtcagcaa ttattgtgct ttgtactaga taattttgcat aaactgtac
69361 tatttttaatt gatgtttatt aaaaacact ttgcatcaga taattttgcat aaactgtac
69421 tcttcattat agttttgggg ggagagaaga ttgcatcaga aaactttatc aaagtacct

69481	agtattataa	aggagtcataa	aaggtacataa	gagaaaaaggt	caagacatttt	ttcaaatgag
69541	ggaaaataaa	caggatttat	cactagtataa	ctctgctctaa	aaagattcaa	gggaagcttt
69601	ttaaaaagaa	gggaagttat	acgagaagga	aacttagaat	ggcagaagata	aaagaagcat
69661	aagtataag	gtaaatataa	taagacttct	ttgaggtttt	aaaaattcata	tttgtatttt
69721	gaagaaataa	aatttaacgtt	gttgatgtgt	attctctgtg	gaggtatatac	aggttttttt
69781	tgtgttctct	gtttctgttt	ttttaaggtg	aagtcctgtg	caccacgaag	tttgtgcagt
69841	ctctgtgata	tggtctcactg	cagcttcacc	ctgggttcag	gtgactcttc	cacttcagcc
69901	ctctgataa	ctgggactac	aggcatgtac	taccacgtcc	tttgtatttt	ttttttcttc
69961	tttttttata	gatggggtct	cactctgttg	cccgagctaa	ttgcataatc	ctaacctcaa
70021	gattattaga	ctatacagta	gactataaga	aattaaagt	ttgtatcatc	ctcagagcga
70081	cacagatttt	tcctataaaa	actacagatt	attttcaaaa	gcattataaa	attaaaggtg
70141	aatacataaa	aagatttcagg	taaccacagga	acgagaaaaa	gaataaaaaa	atagtaaaac
70201	taaaatttat	tttatttcac	taatacagca	ataattaaat	gtcaatgatc	taagctgatt
70261	taaaatctaaa	aactggataa	gaagctgtgg	cttgaatta	aaaaatagac	tttattcacc
70321	agacctacaa	tgtgtttttt	tcatatgtta	agaaaaagtt	aataactctg	ttctctattt
70381	cagttctcag	ggcaccacta	tcagtttaacc	ctaaagattt	ctgtggttga	aaccaatttg
70441	attaccattct	tgctctgctg	ccagttatgg	taaccacaa	ctcttttgtt	ctggatgcta
70501	aaactgtcca	ctgtgtgcct	aactacatcc	taactccctg	ctcttctatt	gaatttaagc
70561	agccaatttt	ttttttttag	atggagtttc	actctgtctc	aggtctgagt	gcaatggcac
70621	gactctggct	caccgcaacc	tcagcttctc	gggttaagt	gattctctgt	ctctcagctc
70681	ccgaatagct	gggattacag	gcaggcacca	ccatgctctg	ctaaatttgt	atttttagta
70741	gagacgggtg	tttccatgtt	tggtcaggct	ggtctggaac	ttccaaactc	aggtgatcca
70801	cccgctcagc	ctctccaaa	tgctgggatt	acagggatga	gccactgcac	ccggccagga
70861	ggcaatttca	atgggagtat	ctcttactat	cattcctgac	ctgtgatatt	cgccattagg
70921	gcttttccag	aatggtagct	ctgctcagat	atttttttta	aaagcttaat	gagagaggtg
70981	tcctcttgat	ctttccagag	aagataggat	gtgagtgagt	gatttccact	taataatcta
71041	accccaaaat	ttctatgtcc	ttgataccat	tcctctgata	ccaagggaag	gtattctggc
71101	atttactctt	gattttaaat	aagtaaaccc	tggtctctgg	catgtatgaa	caaaccaagt
71161	gactacaagt	gataagtaac	ttttaccata	gactgtcaat	taccttggca	cttgaaacat
71221	tattctcttt	taatccagtc	agatcagaga	acaaattcca	gatttccact	atgaagattc
71281	tggtgtctctg	ctaccacttc	tggttagcaaa	acatgtcagt	gtgaggtaga	gtcataacag
71341	gtgagccatg	aaaaaaaaaa	tgctactaca	aacatctacg	gaataaagaa	tttgatcgaa
71401	gaattagcat	ttacacaaaa	tatgggagga	gttaggggaa	tgaagaagatg	aaaggagaaa
71461	gaggtataag	aaaaacactaa	ccacatcacc	ctaaagccct	tattgtctaa	gaaggttatt
71521	gtagtgtccg	ttaagtccag	gtcagccaag	gcgtgcgtac	cccttggccc	cacttggaaa
71581	gctgtgtgga	tatgactgtt	ggatttagggg	gttctcaaa	ccaaaattga	tttatcagca
71641	atgaccttcc	tgctgtggtc	ccagctttga	gaggaccaa	aacgacggat	gtgtcttgca
71701	ttccctgttc	ctcttttctt	cttttaccca	tatctaaagt	gttttaagag	gaaaaatcct
71761	taggttgaag	ccatattctc	tctattcttg	aaggaccaca	gactctacgc	ccctgttgaa
71821	attttctgga	agcctatggg	aaaggaggaa	tgaaaagctt	cggtgatctc	ttgaaatact
71881	gtgctaatct	tactatcgta	actggtccag	taaggagaag	gggtatctga	atttcaacat
71941	gggggctggg	caaggttaacc	agtgaagttg	ggagataaat	agaaactgtg	actgaagacc
72001	ctctcaggtg	gcttgatgat	gcattgtgat	catggcttct	acttcaagat	tggtgtcgga
72061	atagatactc	ttatacttgt	aattccccta	tacacaaatg	ccagcaagaa	tgctccattt
72121	gagacatgct	taaatgtaga	gaggtagtag	tatgacatgt	aaaaacactc	ccaactgttc
72181	tcctgtctctc	ctgcttatgt	ggtacatggc	cactgactac	agttattaac	tttaactgtg
72241	ccctgtgctt	ctgaaaaaca	gtgacagtta	agaaatccct	ccctctgtgt	accatttcag
72301	aaatgtctta	ttgcaaaaga	ccactagaac	tgtagtaaac	accttcatct	accatgatga
72361	aagccagact	cagacccctc	aaatcccatt	ctttgcctca	aaataaattg	ttgaactatt
72421	tgatccacta	gactaatctg	gacaaaatgc	cccttaactt	gacccaattt	tatgactctc
72481	ctctccccgt	aagccactgg	actttgatcc	accttcagcc	tgagccagca	ctgagctgtg
72541	gaattagtcc	tccttttccc	ctcctgaaaa	ttgactagcc	accaggaagg	actctccctg
72601	ctSagctgtc	ccattaacgt	cactgtctcc	tctctccacc	gttagttttt	ttctagcctg
72661	tttattctct	ctctgaagag	aaaaagagaa	aaagctattt	ctatctgatt	tttgaagtat
72721	gtcagatctc	tacctttgaa	gcattctctc	tattgcaata	gtctccctga	cccatattga
72781	atagtctctc	tctccctggt	tgattaccac	cattcaggct	gctatataa	gcacaaatacc
72841	ataagctgat	agcttataga	caatataaat	ttacttccca	cagttctgaa	gcccggaagg
72901	tcaaaagata	aagtgctggc	aggctctggg	tctggtgggg	gacactcttc	ctgactctca
72961	catgttgata	gaagcaaggg	aaactcttgg	gcctctttag	agagcataat	cccactcatg
73021	agccctttac	ctctctgacc	tagtcaattt	tcaaaagctc	gaccacatac	cattcgtatg
73081	ttagagattc	aacgtacaaa	tttggtgggt	tgaggaggaat	acaagcatct	agagcagatg
73141	actattttaa	taatcatttc	aaatagagtc	attctatacc	tagtctgata	ttttgtttgt
73201	tttacttgac	ctcattcact	tggttaggaa	aaaaactcca	tgagactcca	ttgatagccc
73261	caggggataa	gttggggcaa	aaagagcaga	ggcactcatt	tactgtgag	taagaaattc

```

73321 gtttcagatc accttgtgtg caattcagga taaaattaat aaagatgaat agtaaaaaac
73381 caataactaa agataaaaact caggagtcaa aagaagaat cagaagaat agagcataat
73441 tcagaatgaa atgaaaacaa actatgcata gaactctgtg gtattcgccg aaagacgatc
73501 taggagatga atgcaagtat tacctttctt gcttgatata ggataaaaaag ttgaaaaatt
73561 caacagctgaa tatctaacctg aaagagaaaag aaagaaaataa taaaggtaac agacgaattt
73621 aagaaaacaa ccaatactgta gaaaaaacaa agcccaaatg tgagatttcaa aaagagattg
73681 gctgaaaaaa gagttttgtac tatgtattga atgtatttaa aatttaaatg gctcttttaa
73741 tccctcaaac ttaactccttt aaacttcaaa aaagttttaa aagtgctaa ccagctgaact
73801 cacacctgta gtcccagcac ttgtggaggc aaaggtgggt ggtatcacctg aggtcgaggg
73861 ttcaagacaa gctggcccaa catggaaaaa cccgctctct actaaaaata caaaaaaact
73921 gccagcRtg gtggtggggc cctgtaatcc aatccagact actcgggagg ctaaggcagg
73981 aqaatcgata cagcctggga ggcggagggt gcagtgagcc gagatcacgc cattgcactc
74041 cagcctgggc aacaagggca aaactccgtc tcaaaaaaaa agttttaatc ataatttaat
74101 aaatttaaat agataaacaa ggtctcaact ttgtcttcca ggctcgactg aagtgacaca
74161 atcacagttc attgcagttg tgacctcctg ggctcaagct atctcccac ctcagctccc
74221 cgagtagttg gtagggcagg ctattccac cagctaattt ttgtattttt agtagagattg
74281 gggttgtgca gtgttgccca ggcttatctc aaactcctgg gctcaaacga tctgcccatt
74341 tgcactcccc aaagtgcagg gattacatgc atgagccacc acaccogatt tgcctcaata
74401 ctgaaaaact ttttttgaa acggagtttc gcgcttgtcg cccgggctgg agtgccgggg
74461 cgtgatctcg gctcactgca acctcgtctc cccgggtcca agtgcctctt atgctcgacg
74521 ctctctgata gctgggatta caggcacctg tcaccacgccc tggtctaatt ttgtattttt
74581 cgtagagca ggtttttacc atgtttacca ggcctagtctt gaattcttga ccttaggtgc
74641 tccgcctgcc cctggccctcc ctaagtgtcg gattacaggg cgtgagcaac tgcgcccagg
74701 tgtcaaaaact ctttaattag ccggtttggg agtgcagtc tgagatttcca agtatacagt
74761 tgggagagct gtttaagcct gggaggtaca ggctcggtg agccatgact atgacacacc
74821 acaccagcct ggtgacagac ttgttgacag accctgtctc aaacaaaata caaacacag
74881 cgagtataaa aaacttctag ccaccaagcc tagagagttt tatataagag aatagtgatt
74941 tgtgagtttt gctaaatatg cggtttcttt ttattccagt ttatacacga aagtataaag
75001 ctctttgcga aattttacag taacagtacc aaatttaaaa ataactaaat ctcttttttt
75061 tttttgttaa taaggtgatt tgttatttct ttttttttta attatacttt aagttttagg
75121 tgcatctgtc acaatgtgca ggttagttac atatgtatac atgtgccaat cctggtgtct
75181 gcaccogtta actcgtcatt tagcattagg tatactctcc aatgcctatcc cctccccttc
75241 cccccacccc acaacagctcc ccagagtggt atgttcccc ctctgtgtcc atgtgttcc
75301 actgttgtaa tccactctat gagtgagaac atgcgggtgt ttggtttttg tcttgcctat
75361 agtttgcctga gaatgatgat ttccaaattc atccatgtcc ctacaaaagga catgaactca
75421 ccatgtttta tggctgata gtattccatg gtgtatatgt gccacatttt cttaactcag
75481 tctatcattg ttggacattt gggtagtttc caagtctttg ctattgtgaa tagtgccga
75541 atagacatac gtgtgcatgt gtctttatag cagcatgatt tatagtcttt aggggtgata
75601 cccagtagtg gtaggtgtg gtcaaatggt attttagttt ctagatcccc gaggaactca
75661 cacactgact tccacaatgg ttaaaactagt ttacagtcac accaatagtg taaactattt
75721 cctatttctc aacatcctct ccagcacctg ttgtttcctg actctttaat gattgccatt
75781 ctaactgtgtg tgagatgata tccattgtgt gttttgattt gcaatttctc gaaggcag
75841 gatgatgagc attttttcag tgccttttgg ctgcataaat gtcttctttt gaagagtgtc
75901 tgttcatact ctttggccac ttttttagtg ggttttttgg tttttctgtc taaaatttct
75961 gggagtccact gtatagttct gatattagcc ctgttcaccc ttgttgcaga ttagtaggtt
76021 tctccactc tgtaggtttgc ctgttcaccc ctgtgttagt ttgttttctg ctgcagagc
76081 ttttagtttt aattagatcc catttgtcaa ttttggcttt tatgtccatt gcttttgggt
76141 ttccagacat gaagtcctgt cccatgccta tgctctgaat ttgtcttctc aggtatttct
76201 ctagggtttt tatgttttta ggtctaacgt ttaagtcttt aatccatctt gaatttaatt
76261 ttgtataaag gttaaaagag ggtaccagtt tcaagcttct acatagtcct agccagttt
76321 cccagcaccc ttatttaaat aggggaactct ttcccatgtt ctgttttttc tcaggtttgt
76381 aaagatcacg atagtgttag atatgtggcg ttatttctga ggctctgttt gcttccatt
76441 gatttatatc tctgttttgg taccagatcc atgtctgttt ggttactgtt gcttgtagt
76501 atagtttgaa gctcaggtagt gtgatgcctc cagctttgtt cttttgctta aggttgact
76561 tggagatggt ggttcccttt tggttccata tgaaacttaa agtagttttt ttccaaattt
76621 tgaagaaga gcatgtgtag cttgatgggg atggcatgaa atgcataaat gaccttggc
76681 agtatggcca ttttccaaat attgatctct cctacagat agcatgaaat gttcttccat
76741 ttgtttgat cctcttttat ttccctgagc agtggtttgt agtctctctt gaaggagtc
76801 tccagctccc ttgtaagttg gattcctaa gatttcttat cctttgaagc aatgtgaa
76861 gggagttcac tcatgatttg gctcctgtt tgcctgttaa cgggtataa gaatgctgt
76921 gatttttga cattgatatt gtatcctgag acctttgctga agtgccttat cagcttaagg
76981 agatttttgg ctgagacaat ggggttttct atacatacag tcatgtcttc tgcacacag
77041 gacaatttga ctctctcttt tccataatga ataccctta ttctcttacc ctgcataatt
77101 gccctggcca gaactccaa cactatgttg aataggagtg gtgagagagg gactccctgt

```

77161	ctgtgtccag	ttttcaaagg	gaatgcttcc	agtttttggc	cattcagtaa	acatggaaaa
77221	atagcaaaat	agagacagtg	ggatataaca	attgcataata	tcacaataaa	tgtgaigtta
77281	atctcagaaa	tgcattggcg	tctgtccttt	agaaaatgta	ttttttgat	ttttatcaat
77341	ttagaggtgc	ccttttaaaa	aaacataata	caaatgtgaa	tattcaaaa	gcagctgttt
77401	tatttctctc	cttgcgtgaa	catataataa	tgcagctgtt	taaaaattat	gaatcttttg
77461	attgtatcaa	atatggtaat	gtaatttggc	atttccacag	atcaaaatca	gaatcagtat
77521	gatcatctat	cttgcatttc	gagaagcatt	tgaataattt	caaaatccat	tgtgaaaaa
77581	agctcttgga	gaggagaaa	agcaaaagaa	ttttaaaata	taaaaggtat	tgacaaaact
77641	gtagtaacaa	tcttcttgaa	agatcaaaat	tgttcccttt	cagcttggga	ataagacaga
77701	gcttccattt	aagaccacca	catcatctca	tctttaaagt	tggtgattct	accacacat
77761	ctaagcatga	aagaaaataa	aattataaga	taaaaaggaa	agaaaactct	cagtaactaa
77821	agatgatata	attacataga	aaatcaagaa	gattccatgg	ataagcttcc	caaaaaacaa
77881	tatgtaacat	accagcaaca	aatggttaaga	aaacacagtt	gttttttttt	taaaagtacat
77941	tttcatgtac	attaaataat	aaggatttcag	gaatacaacg	aaataacaaa	aaatgtaaaga
78001	tatctagggg	tgcgtataac	aaaaaaatta	taaaagcttca	tttaaaaact	tttttaacaa
78061	ataaagtaga	aagtaatacc	atcttttatg	acacagatca	gattataaaa	tgtgtcagtt
78121	ttcttcaact	ggctcataga	caacacaatt	ttactaaaga	ttccaaatgg	gatttctctg
78181	gaactctgat	atttagggcc	aaaagttaac	aagatattca	tgaggaagga	taacagaatg
78241	gcaaaacttg	ctttaataat	ttgtgcaggg	atagacaaat	actgtacacc	agttctccaa
78301	tagggagatc	aaaaacaaat	tcacacagac	atgaaaaagt	gatttatgac	agatgatgtt
78361	gcagctccct	gggaagaagt	ggaaatgggt	ctagaataat	tgttttctat	atgtatttct
78421	ataatttatt	atgcccattac	attaaagtca	atgcttatgg	agtttgggag	gagtgatcat
78481	aatgttctg	aattagatgt	gatgtttgca	caactctgtg	aatgtactaa	gaacccatga
78541	attgtatatt	tttaaagcat	ggtagtga	taatactctca	catctctatt	aaagtaccaa
78601	atttttaaaa	tcccagcaat	ttgggaggct	gaggcgggcg	gatcatgagg	tcaggagttc
78661	aagaccagtc	tgaccaacat	gttgaaaccc	catctctatt	aaaagtacaa	aaatgtacca
78721	gcagtggtgt	cagtgccctg	taacttaagc	tactcaggtg	gctgaggcag	gaaatctgct
78781	tgtatccctg	agggcgaggc	tgccgtgagc	cgaatgatg	cgcttggcag	tccatccagc
78841	ctggggcaca	gagcgagatt	ccatctcaaa	aaaaaaaata	aagtaaaaaa	tcaatgacgg
78901	ccagcgcgac	tggtctcacac	ctgtaatccc	agcactttgg	gaagcgcagg	tgccagatga
78961	ctcaggtgca	ggagttctgag	accagcctga	ccaacatgga	gaaacccctg	ctctctataa
79021	agtaacaaat	taqccgggtg	tgggtggtga	tgccgtgaat	ccagctact	gcggaggtct
79081	aagcaggata	atcacttgaa	cccggggagg	ggcggtgtgt	gtgagcaaa	aaacgcctat
79141	tgcctccagc	ctcggggcac	aagagcgaaa	ttccatctca	aaaaaaaata	aaaaaaaatc
79201	aatgcccagc	agactgaaaa	cttaacatgc	aaagggcaag	tattgaagac	caaacagata
79261	gcaatatttc	accacaagag	ggcagtttga	ccctgaactt	tgggtgtgtg	ttgtttttag
79321	attgagctct	gctctgttgc	ccaggcttga	gtgcagtgcc	gtgatcttgg	ctcactgcac
79381	ccctccacct	cccgatttca	agtgattacc	ctgcctcagc	ctcccgagta	gctgggaact
79441	caggccagct	ccaccatgaa	cgactaattt	tttgtgtttt	tagtagagac	ggggtttcac
79501	cggtgttagc	aggatggcct	tgaatccctg	accttgtgat	ctgccacact	cagcttccca
79561	aagtgtcggg	attacagggc	tgaactttga	ttttatacag	gaagtggttg	gaggaaatgg
79621	ttggaaagct	gttttcgtga	ctcaaaagac	tactagctct	taactagctt	gagtagatgt
79681	gccacaagtt	aaagttaccc	tggaatccag	tcaaaagctg	tgttatgatg	ctagttttgag
79741	ctagtatctt	ctcggagcca	gtaacatgaa	tatgatgacc	aaagtgttat	ttcgtctctg
79801	acaaaagagt	tatgacacac	acttcgttgt	atgtgtccat	ttaggccac	cgtaaaagag
79861	tgaagcacgt	gcatacaacca	caaaagaaat	tattgtgaca	ctttctctgg	atgataaaat
79921	atcttctatg	tgcgtgtgca	tgtgttgtgt	atgttgacag	gtcacacatc	agaaactgcc
79981	acctcatagc	cgaaggatag	tcctaatatg	aaacgcataa	tgaaggatgt	tttaataata
80041	taggctttgt	tgcttccctg	ccccggcatt	gccactgaat	tgtgtgactt	tgggcaagtt
80101	acctatgttt	cagtgcaact	gtataaatgag	gataatatta	gataatccct	catggtattt
80161	tttttctctt	tttttttaat	tatactttaa	gttctagggt	acatgtgac	aacgtgcagg
80221	tttgtttacat	atgtatacat	atgccaatgt	gggtgtctgc	accocatatt	tcgtcaattt
80281	cattaggtat	attctcttaat	gctatccctc	cccatccccc	ccaccccaag	acaggcccca
80341	gtgtgtgatg	tttcccttcc	tgtgtccaa	gtgtctctct	ttgtcaattc	accctatgag
80401	tgagaacatc	cagtggtttg	ttttctgtcc	ttgcgataat	ttgtctcagaa	tgtgtgttct
80461	cagctctatc	cagtcoccta	caaaggacat	gaactcatcc	ttttttttct	ctgcataatg
80521	ttccatgtgt	tatatgtgcc	acattttctt	aatccagctc	atcactgatg	gacattttgg
80581	ttggttccaa	gtcttttgta	ttgtgaatag	tgccgcaata	aaacatcgtg	tgcatgttgc
80641	tttatagacg	catgatttat	aatccctttg	gtatataacc	aaatattgaa	tgctgggttc
80701	aaatggtatt	tctagtttct	gatccctgag	gaatcaccac	actgtcttcc	acaaaggttg
80761	aactagttta	cagtcocacc	aacagtgtaa	aagtattctc	gtttctccac	tctctctcca
80821	gcacctgttg	tttcttgact	ttttaatgat	tgccatttct	acgggtgtga	tgaaaacttc
80881	tcagagctaa	aagaggaagt	tcgaacccat	cagaaagaag	ctaaaaacct	tgaaaaaaga
80941	ttagatgaat	ggctaaactg	aataaccagt	gtagagaagt	ccttaaatga	ctctgatggg

81001	ctgaaaaatga	tggcacgaga	actecqtgac	aaatgcacaa	gcttcagtag	ccgatttgat
81061	caactggag	aaagggatc	agtgattgaa	gatcaaacga	atgaaaagag	gtgaaaagag
81121	aaagttagag	aaaaaagagt	aaagaagaa	gaataaaagc	tccaagaaat	atggggagat
81181	gtgaaaagac	caaatctatg	tctgattagt	gtacctgaaa	gtgattggag	gtgataaac
81241	aaagtggaaa	acactctgca	ggatattatc	caagagaatt	tcoccaaact	agcaagcgag
81301	gccaaacttc	aaattcagga	aaatacagaga	acaccacaaa	aactctctcc	aaagaagaga
81361	actccaaagc	acataattgt	gagattcacc	aaagtgtaaa	tgaaggaaaa	aatgttaag
81421	gcagccagag	agaaaggtcg	gggtaccac	aaagggagc	ccgtcaaaat	aacagcagat
81481	ctctcgggag	aaactctaca	agccagaaga	gagtgggggc	caaatattcaa	catctctaaa
81541	caaaagaatt	tccaacccag	aaattttacat	ccagccaaac	taagttctcc	aaagtgaaga
81601	aaaaataaat	cctttactga	caagcaaatc	tgagagactt	tgctaccacc	aggcctgcc
81661	tacaagatgc	cttgaaggaa	gcactaaaca	tagaaagpaa	caaccactac	cagccactgc
81721	aaaaaatgac	caaaatgtaa	agaccattcaa	tgctagaagc	aaactgcatt	aactaacgag
81781	caaaatgaac	agctaacatc	ataatgacag	gatcaaatc	ccacataaca	gtattaata
81841	taaatgtaaa	tgggctaaat	gctccaatta	aaagacacag	actggcgaat	tggataaaga
81901	gtcaagacc	actcagtgtc	tgatttcagg	agaccatct	catgtgcaga	ggcacacata
81961	ggctcaaaat	aaagggagg	aggaagatct	accaagcaa	tggaaaacaa	aaaaagtag
82021	gggttgcaat	cctatctctg	ataaaacaga	atataacaa	acaaagtag	aaagaagaga
82081	ggcattata	taattgttaa	tggatcaatt	caacaagaag	agctaatct	cctataata
82141	tatgcacca	atacaggagc	accagatctc	ataaagcaag	tccttgagga	cttacaagata
82201	gacttagact	ccccacaaat	ataattggga	gactttaaca	ccccactgtc	aacattagac
82261	agatcaagga	gcagaaaggt	taacaaggat	atccaggact	tcacctctgc	accaagggga
82321	cctaatagac	atctacagaa	ctctccacca	caaatcaaca	gaatatacat	tcttctcagc
82381	accacatgct	actatttcca	aaattgagca	catagttgga	agtaaaagac	ctctcagaca
82441	atgcagaaag	acagaaataa	caaaactgtct	ctcagaccac	agtgcaatca	aattagaact
82501	caggatttaag	aaactcactc	aaaaccactc	aactacatgg	aaactgaaca	ccctctctct
82561	gaatgactac	tgggtacata	acgaaatgaa	ggcagaaata	aaaggtgttt	tgaagaccaa
82621	tgaagacaaa	gacacacagt	accagatctc	ctgggacaca	tttcaagaag	gtgtatgagg
82681	gaattttata	gcactaaatg	cccacaagag	aaagcaggaa	agatctaaaa	ttagcacctc
82741	aacatcaaca	ttaaaagaac	tagagaaagca	agagcaaaac	catctcaaaag	ctagcagaag
82801	gcagaaataa	actaagatca	gagcagaact	gaaggaaata	gaacacaaaa	aaagccctca
82861	aaaaatcaat	gaatccagga	gctgggtttt	tgaaaaagac	acacaaaagt	ataagcttca
82921	agcaagtcta	ataaagaaga	aaagagagaa	gaatcaaaac	gatcgaaata	aaaaagtata
82981	aggggatctc	accaccgact	ccacagaaat	acaaactacc	catcagaact	actgtaaata
83041	cctctgtgca	aataaaactag	aaaatctaga	agaaatggat	aaattcctgc	acacatacac
83101	ctccccaaga	ctaaatcagg	agaaagttga	atccctgaat	agaccaataa	caggctctga
83161	aattgaagca	ataatttaata	gcctaccaat	caaaaaaaga	cctcatgtga	tttttatgga
83221	atgtaaatca	gtttatagat	gagcaggagt	ctagtatcat	cctttaaagac	atactctggt
83281	tttatagagt	cagaccgggtg	aattaaagtag	gcacttgatg	gggatcacag	tgccctgcaac
83341	ttttgggtct	ttaaagtaga	ctttaacgtc	taccatagtc	ccaggcaata	ttgttttttaa
83401	tttaacaatat	aattttttta	catttttaca	attattttata	ataattgtct	tttatgtaga
83461	atattgtcoa	tgccgtattc	ttttaaaatt	tactataattg	gctgagacag	gtatagagag
83521	acagttttag	agttcccaact	tggtcttccc	cacttccgta	gctcttaaac	catgtgttaa
83581	gaaccaggtg	tgggggttgt	gctaactaca	aagtcagtag	ccagggaata	aaactcagat
83641	gggtcatgaa	accagtgaac	ctactgtgag	tgggattctt	gcaaggagacc	catgtatcat
83701	ctgggcaatga	taaccctttt	cccacaacaag	gtggccatga	taatgcccc	ttgaattaat
83761	ctctgattaa	ttcttagtag	ttaatgtcaa	acacagactc	gccaccaaca	aactctctca
83821	aatgccttgg	cttgtctctt	tccttgtgat	acagtaacct	ttttaaagtg	ccacaggtct
83881	ctttgggata	ggtatgagag	aatcatgtta	aacacttgca	ttataggggt	tttctcacgc
83941	aagatctcca	ttgttgttca	gtcaattgggt	tctgaactcg	aaaaaggtct	caactcaggt
84001	aacagggag	acatcatgac	tttttgttgg	gaagtgctct	caggatttta	gtcatctatc
84061	tttgttgtgt	atcatatggc	atgtactctt	tggtctggcg	ttctttcaat	ttgatcaagt
84121	atttttagat	ttacacagat	tgtagtagca	tcgatctttt	tatttttgaa	tagcatccca
84181	tttaattgat	tcatttgatt	gcaatgtctc	ctgtctctct	tgcccttagc	gactcaatgt
84241	ctcttaactc	tcatagtctc	ccaggggtcg	gaaactccat	tacttgctac	tcagactttg
84301	ctgttatatc	agtaattaa	cctatttttg	ttctgtgggt	ttagatttgc	ccacaggcga
84361	cttttgtttg	gggactctgt	tattcccaat	gccatcatgt	agcgtagttt	ggtaaccaac
84421	taaccataga	gacttagcca	ctgggtgtcta	gaaccagctc	tcacaagatt	gatttatctc
84481	catcaactct	cacaagattg	ttacattttt	tgaagtttta	tgaaccaggt	gttaccctat
84541	agtgtgaatg	taatatgaca	ggagaatgat	ccatctttta	gtcgatatac	ttccctctat
84601	ccaacattat	tctccttgac	tgcattttca	tgctccagtc	tcagtgtgat	tttctcagct
84661	ctgtccagat	cactttaact	aggacctgca	gcgccttttc	ctccataaca	agccagacac
84721	ccccaagctg	tgttatatca	tgattgtacc	taatttttta	ttgggtagcc	agaattgtga
84781	tggggggttg	catctactctg	ttctccactt	gctataaaga	aatcactctg	ataggtaaat

84841	tgataaaaaga	ggtttaattg	gctcacgttt	ctgcaggctg	tacagggaagc	acagtgctaa
84901	caagtgtctg	gcttctgggg	agggctcagg	aaacttacia	ccagtggcaga	aggtggaagg
84961	ggagcagcga	tatcacatgg	agcagcagcg	agcagaagaaa	tgagggggag	atgcttttaa
85021	atgaccagat	cttacaagag	ctcacctact	gttgctgatg	cagtaaccaag	acgaggtgtg
85081	taagccattc	atgagaaatc	tgccaccatg	atccaatcac	ctctgaccat	ggctaccctc
85141	caaacattgg	gtttttttat	atgagatttg	gggtggggaca	cacatccaaa	gtatctacg
85201	tgtgtattct	gggcagaggt	gggtggggtg	gggtggggcat	ttatcttgca	aaagttagaga
85261	cttggcatca	tcttcacagg	aatgggtaga	gcttaaccttt	aaagagagaga	aggaggaagg
85321	ctcgtctgtc	cagcccgagg	gaatcgagg	aaagtgcatt	caagattttc	agcatattaa
85381	ccaagagdac	ccattccaa	tcttaagatc	acagctcttc	ccatcaatac	ctccctctgac
85441	cttgatgaag	ccaactgtgc	tgagtttgaga	atcatatatt	tcttagagctt	gttccactact
85501	cttgatgaat	aaggctgaat	ctgatctcca	gttttttctg	cttccagact	accaagagatt
85561	agcatatgct	gccaaagagg	ccctctagct	ttcacacctt	gccctcagtc	gatgattaat
85621	acccttagct	tctcatttcc	ccccaaagt	atctattgat	agcaatttagc	atcatcatata
85681	ttccactttt	ttgttttaac	gattttatga	agtttaattg	atgacataaat	actcatattg
85741	gtgtttgatg	tatgtgtgtg	cagtgacaaa	tttgctaagt	ttttgtttgt	ttgttttttt
85801	gagacagggt	ctcgcttttg	cacacaggct	ggagtgcaat	ggcacagctga	tagctcactg
85861	ccactccagc	cttccacatt	tggttaagtt	tgacatrtat	ctatgaacac	atgtacaaaa
85921	tcaagaaagc	aaacatttca	tcaaccctaa	aaagtctctc	ctgctctttt	tgactttttc
85981	cttccctcta	tcccagtgta	actcctgatt	tgctttctga	taggttagttg	tgattttttt
86041	agaattttat	ataaatggaa	taatacagca	tgattctctt	tttttttttt	gtctggctac
86101	tttcttaaat	tataatgatt	ttgagattta	tgatctcgt	ttgtgtatca	actcttttat
86161	tggttgaaat	tgatcccat	ttaatagatc	tatcaaatgt	gtttatccat	tgatctgttg
86221	atggacattg	gggtctattc	caggtttggg	ctattatgaa	cttccagctc	atgaattatt
86281	atatacaagt	ctttgtgttg	acatgtttcc	atattttctg	aataaatatg	tagaagtggg
86341	atgtctggga	tctttggcaa	aaagtgttta	agttttcaag	aaagtggcca	ttctttttcc
86401	agtcattctg	ctcttttgca	ttctcatcag	cagtgtagga	gaggtccagc	tgccccacac
86461	cttcaccagc	agtggggcaa	gtcttttaac	tttagccatt	ctagtgggtg	ttatagtggta
86521	actcatttga	gttttttcat	ttcattttta	aatgtaggta	caattgacat	gcaatgaact
86581	gcaccttaat	aaactgtaca	attgtatgag	tttgacatat	gtataacca	atgaacacca
86641	cacaatccag	atgataaata	tattcatcac	cttcaaaagt	tttctcatac	cccatgtaaa
86701	ttctatcttc	tctccctccc	ttgctcctga	ttccctaccc	cagataacca	ctaactcttc
86761	ttctgtcatt	agtttgcatt	ttctagagtt	ttatacaagt	ggagatctaca	gtatgtactt
86821	tatatgtatc	aataggtcat	acctttttgt	tactacatct	tttgttactg	tgacatattt
86881	cactgtatgt	ctaaaccatt	attagtttac	ctgttgacag	acatttgggt	gttttctaat
86941	ttttgcccga	cacaaataaa	gtggctatga	acattttacat	acagtgattt	tgctggactt
87001	aagctttaat	ttctcttagg	taaataacct	aggagtagaa	tgctgtggtc	atataatagg
87061	tgtatatagt	atttaacttt	ttttataaac	actgtaaaac	tgttttcatg	tggtgtttatg
87121	gaataaaaaa	tgagaaacaa	aataaaaaca	atgtatacca	aaacatgttt	aatgtttcac
87181	cttcaagaaa	agggcgcact	ttacaatgag	tgagaaaaat	tatatccaca	ttttgaaaaa
87241	aactccaaga	taaaaggcat	tttgacaaca	gatattttga	aatattttcca	tcgttacaag
87301	agttttattg	caaaaatgat	gtatgtccat	ctataaaact	ctcctatctc	ttgttataaa
87361	aaattttact	gtctattttt	taaatatttt	caaaaagtta	gtttttaaacc	tatttgttta
87421	aaatatgaat	atgcaaaaac	ttttattgct	tccaagagta	gttttttgac	atctgagaaa
87481	atataaaatt	attactttaa	caaaaactct	ccctttttagc	tcctcttttt	tgttttgttt
87541	ttataactgg	acagataata	attgtacata	ttcatggggt	acatagcgat	ttgttcaaat
87601	atatactgtg	gtgattagat	caggytaatt	aaacatatct	tttaatttgt	gaatggagtg
87661	aaaaaattaa	actctgattt	agtaagtgaa	gctaactctt	ccatttgatt	ccatcacttc
87721	ctgcaggatg	tttttttagt	ttgagagcca	ttaacaccaa	atattggaat	aaactcaact
87781	tataaccaga	ccatcatatc	attttatctc	ataaaatgaa	cgaaagtatg	ctgtcacttc
87841	tggaacacaa	actgaacagt	attgtttact	atgtaaaaaa	aatgtagttt	tcaagtgaaa
87901	attggaattt	tggaaaaact	gtatcctccc	tatgagcttg	acagcttccc	aatattttaga
87961	gactttttat	atataattat	ggtgagacca	gtgataatat	taacaattgt	attttttgaa
88021	atattgtcta	atgaaaccca	tcaaatccaa	cccaaatttt	ataaaaaata	agcatataaa
88081	aaaaacttta	ttggaggggc	atggtggctc	atgcctataa	tccagaccat	ttggggagcc
88141	gaggcagcca	gattacttga	ggccagaagt	tcaagaccac	cttggtccaa	atctgtgaaa
88201	ctccatctct	acttaaacaa	aaaaacagaa	ttcttgcaat	ataggtgacc	ctcataaact
88261	acataactct	cttgagagta	aaagaggtgc	atttttttgt	agagttttct	agctactttta
88321	catataattat	atttataatt	ttcagatccc	aaattggcga	aaagtggaga	attttttcta
88381	aaattcaattt	tcttctgtgc	agagccaaat	atgtaagaaa	actcttagat	accatggtta
88441	tctccataat	cttaaaaatta	ttgttctcag	cacaatgaag	aaattattga	gaacagagctc
88501	cttccaaaga	aaaaaaaatc	caataagaga	tttaatatgt	attgtaattt	tgaagaagag
88561	ttaagttgca	ttctcacata	gaatgctctc	ctcagacactg	gcttattctca	aatgcytgac
88621	caagtaatca	aaaggtacag	taaatgccta	ttcttctcta	gaacactgct	ggacagagatt

```

88681      tccccctaac      cctggetaag      gctctcttta      tttattagat      ggattaaact      taattccctt
88741      gatttgtact      tattctaata      acagttgatac      cataggaagt      cagaagaagt      ttttcagact
88801      tgagatttaa      gtgcaaatgtc      agcttttttt      gttttttttt      cttttttcag      tttyggtcaa
88861      gcaagaacaag      ggaattgtgt      ccaagttagag      cacactatgc      ccaactaggg      acataataga
88921      tgactcagca      ttgcatttta      gtctcatgac      ttgtctggga      atcagatgca      gctatcgaga
88981      ataccagag      cattggcttt      aggttttagta      gccatgagat      aggttagga      aggcagaagt
89041      ctactacaag      aaactcctga      ttattgtctgc      aagaacttct      tggtaaaagc      attcttgaca
89101      atactacctt      gaaaaactct      ttaactatgt      gaaaaactga      aaaaactagg      gtgagagac
89161      aggcctctggg      ttgaaacctt      ggctatacca      ccaacttacac      atcatgtgac      cgtgggttaa
89221      tggctatcct      accctgtgtc      tcagttaacca      tatatacaaa      atacaactaa      ctaactgcta
89281      aatacttact      aaactcctta      cattattat      gaataaaac      cacatgatcc      tatgtggctt
89341      atattaataa      taaataaaaa      ggatttagtc      tggtaactaa      tgaccagtaa      atgttattta
89401      taattgggtt      attttattat      ttatttattt      agacagagtt      tcactcccat      tgccatgggt
89461      ggagtgcagt      ggcaactcagg      tgattctcct      gccctcagct      cccaagttagc      agggacacaa
89521      ggtacacacc      catgccaccg      taattttttt      tttttttttt      tgagattggag      tctcaactct
89581      tcaccaggct      ggagtgcagt      ggcatagtct      cggctcactg      caactccac      cctctgggtt
89641      caagcaattc      tctctgcctca      gccctccaag      tagctgggag      ctacaggctgt      cggccacact
89701      ttcactgaat      ttttgtattt      ttagtagaga      cggggtttca      ccatattgtc      cggatgggtt
89761      tcaattctct      gatcttgtga      tctgccacc      ttggcctccc      aaagtctggg      gataaattgc
89821      gtgagcgacc      aacctgggcc      tgcataattt      taattttttt      ttgtagacaa      cgggtttcac
89881      atgtgtccca      ggctgctctt      gaactcctga      gctcaagtga      tctgcctgcc      ttggcctccc
89941      aaagtgcgga      gattacaggt      gtgagtcacc      atggctgacc      agtttgtctt      attctttaag
90001      cctcaacaaa      tccctgtagt      taattgtggg      taggtatctg      ctatgggttg      aatgtccact
90061      ccaaaaagta      tgttgaattt      taattgccat      tgtaacagta      ctaagaggtg      gaacctttaa
90121      gaggtgatta      ggccatgaga      gctctcataa      atgtattaac      gacattattg      tgggagtggt
90181      ttcatattga      caggacaggg      ttctcataaa      aaggacaggt      agccacccc      aactgaggtc
90241      ctctgactct      caaggctctc      tgcttttcca      cctcttgcta      tgggattgtca      taacaagaag
90301      accctcactc      gattgctgga      ccttgatctt      gaactttccc      acctgcagac      gtatgaggac
90361      ataaatttct      attcctctaa      aattaccagg      tctgtagtat      tctgtatatg      cagcacaaat
90421      cagactaaga      cagtatccct      aagggttaaa      tgaatttatt      tctgagatga      aactgaggtc
90481      cagaagttaa      ttgtccattg      acccagagcc      tgtgtgtggc      agtttctgtg      attccaaaac
90541      aggcactctc      caatatgtca      tctcaaatct      acaaaattca      ttatgtctt      gtaagtgtag
90601      aaatcagaaa      catcatacaa      gctcagtcct      ttacgtgggt      ttgactctgc      ccaaatttaa
90661      aagatttata      gaattataaa      ataagtgtat      gaggtgggct      gcaatggctc      atgctgttaa
90721      tccagcactc      ttggaggacc      gaggtgggtg      gatcacctga      ggtcaggaggt      ttgagaccag
90781      cctggccaac      atggcaaaac      cctgtctcca      ctaaaaatac      aaaaattagc      cagtagtgtt
90841      ggtgcatgcc      tgtagtccca      gctactcagg      aggtcaggag      agggagaatt      ctgaaacccc
90901      ggaagcggag      gttgcagtga      gccaaagatc      caccactgca      cccagctctg      cggtaacaga
90961      cgagactctg      tctcaaaaaa      aaaaaataaa      ttaattaaaa      agtgtgtgat      attgaacttt
91021      gactctctct      gaacaaagcc      tgcattattt      aattgttgag      ttattttaat      tttctgttga
91081      ctatattttg      ttttgttgcc      ttaatatgat      aataggacag      agacagattt      gtgagagatt
91141      cagcacagct      gctgaaaaaa      ttgaggatgc      tagaaaagtt      cttttatata      ttaactctat
91201      cagcagatat      ttattgcagc      ccttatgtga      tcaggacact      gctgtgatca      agcaatggat
91261      acatcagaca      aggttctctg      tcttatagaa      atgacattat      agagagagga      gaagaacatt
91321      tattccatca      ataaatcat      aaaaaagaca      atttcagata      gtgtaagttaa      cagctaagaa
91381      aacaataaaa      caggatgatt      ggatcacagga      gatggggaag      gatgggtgag      agtggggaga
91441      gcttagatat      agttaaattg      aaaggcttgt      ctttgagggt      ggcatatag      ctcaagtact
91501      cttttttttg      agtcagagtc      tccctctgtt      gccaggctgt      gagtcagtg      gcaYgatctt
91561      ggctctctgc      aaacctgtcc      tccctgggtc      aagcgattct      catgcctcag      ctgggattac
91621      aggcgtgtgc      caccatgcc      aggttaattt      tgtatttttt      ttgtagaca      gggtttcgcc
91681      gtgttagaca      gcctggctc      gaactcctga      ccgcaaatga      tcagcccaac      tcagcctccc
91741      aaagtgcgga      gattcacagg      atgagccact      gcacctggcc      ctcaataact      tgatcaattg

```

[0241] The following is a human cDNA sequence for TM7SF3.

TM7SF3 cDNA Sequence (SEQ ID NO: 2)

NM_016551 Homo sapiens transmembrane 7 superfamily member 3 (TM7SF3), mRNA

1 ccagcctctgg cgtggggccca gcccgccca gccagcaatg gggttcctgc agctgctggt

```

61  cgtgacgggtg  ctggcatccg  aacacccgggt  ggctgggtgca  gccgaggtct  tcgggaattc
121  cagcgaggggt  cttattgaat  ttctctgggg  gaaatttaga  tacttcgagc  tcaattaggcc
181  cttccagag  gaagctattt  tgcattgatat  ttcaagcaat  gtgacttttc  ttatttttcc
241  aatacaactca  cagtatcaga  atacaactgt  ttctttttct  ccgactctcc  ttccaatttc
301  ctgcggaacaa  ggcactgcc  gtggactggct  ttctatcctt  agaccagagc  agagtacatg
361  cacttggtagc  ttggggactt  caggcataca  ggcctgtccag  aatatggcta  tccactcttc
421  ctactcagaa  agagatccct  tccctggagg  ctgtaatttg  gagttcgatt  tagatatgga
481  tcccaaacatt  tacttggagt  ataatttctt  tgaacagact  atcaagtttg  cccgaccaa
541  cctagggtact  gcgagaggcg  tagatccccc  accatgtgac  gctgggacag  accagagctc
601  caggtggagg  ttgcagtatg  atgtctatca  gtaatttctg  cctggaatg  accctcagta
661  ggagatgttg  ctgaagcatc  tgcagaggat  ggtcagtggt  ccccgagtg  agggcagctg
721  tctcaaggtg  gttaccctaa  cagctaata  taagacaagt  gtttctctct  cctccctccc
781  gggacaagggt  gtcataata  atgtctattg  ttgggacccg  ttcttaata  catctgtctg
841  ctacatttct  gtcacacat  acgcttgca  ctttgaggca  ggagagggt  gttgtgcttc
901  cctagaagaa  gtgtcttcca  aagtgttct  cactcttttt  gccctgctgt  gtttcttcat
961  ttgtttcttt  ggacacagat  tctggaaaac  agaattattc  ttcataggct  ttaactcat
1021  gggatttctt  ttttatata  tgattacaag  actgacacct  atcaagtatg  atgtgaatct
1081  gatttgaca  cgtgtcactg  gaagcgtcgg  tgggaattgt  ttggtaggcg  ttggtggcg
1141  atttggaaat  ctctcgatct  ccatgctctg  tgttggacta  gtgctgggtg  tctcatcttc
1201  ctgagtgact  tcttttactc  cactgggaaa  cctaagaatt  ttctatgagt  atggtgtatt
1261  ctgggtcact  tctcttgca  tagctatcct  cattccagta  gttttcatg  gctgctatg
1321  aatactgaac  atactgaatt  gtggagtc  ttgctcctat  tcggtgtctt  tagccattga
1381  cagttactgg  tccaacagcc  ttctctacat  cactttgaac  gtactcaaga  gagcgctcaa
1441  acagagtttc  cacagagctt  tcacaagtgt  gccctttcaa  actaatgact  tcaattattc
1501  ggcagtatgg  ggcactgctg  ctgtaagtgg  aattacgtta  cagattcgaa  gagagagagg
1561  acgaccgttc  tccctcccc  accatacaa  gttatggaa  caagagagag  agcgccaggt
1621  gacaaaact  ctggacccta  gctaccacat  tcttccatg  agagagaggc  tctatggccc
1681  attaacccag  attaaaggcg  tcttccagaa  ggagacacca  gctggcgaga  gaacgctttt
1741  gctctgttag  atgcccaggg  gcttggctag  tgtgctcag  ctttggagt  catgctgtg
1801  gtggttcaac  agtctctggt  gcaagtctaa  taagagatca  ggcataata  tctgttcttt
1861  gcaataatt  atggtgccct  tattgatata  tggtaagggt  gtactagggt  attagagatg
1921  ttgtaagaga  atgagaaga  tgacaaaag  gttgtgtgta  gggagctctt  tctttatttc
1981  caaatacttg  agaaattacc  ttttggttta  caaatctatg  atcaacttat  tccatttaaa
2041  agatcattta  aaaaaattaa  aaactgcata  aaaaaaaaaa  aaactgggtg  tcttttttat
2101  aacccttga  acaagctctc  tcactgagc  ctgtctaaac  tttcggagg  agtttatatt
2161  tgagctctta  tctgtgacag  tatttgaga  tttagggatt  tgatacttag  ccttttgat
2221  tttagaatac  aaaaagagaa  gcaagccaga  catggtggct  cacactgta  atcccaatc
2281  tgggaagcca  aggtgggagt  atcgcttgag  cccaggagtt  tgagaccgac  atgggcaaca
2341  tgacaagacc  ccatctctac  aaaaaaatt  aaaaaattag  ccaggcatg  tggcacatgc
2401  ctactccagc  ctcccaagga  gactgagatg  ggagatccc  tggagccctg  aagattgagc
2461  ctacagtgg  ccttgattgt  gtcactgcac  tccagcttgg  gtgacagaga  cctgtctctg
2521  agaatt

```

[0242] The following is a human polypeptide sequences for TM7SF3.

TM7SF3 Amino Acid Sequence (SEQ ID NO: 3)

NP_057635 transmembrane 7 superfamily member 3; seven transmembrane protein TM7SF3 [Homo sapiens]

MGLFQLLVAVLASEHRVAGAAEVFGNSSEGLIEFSVGKFRYFELNRPFPPEEAILHDISS
NVTFLFIQHSQYQNTTVSFPTLLSNSSETGTASGLVFIILRPQSTCTWYLGTSIGIQPVQN
MAILLSYRSDPVPVPGCNLEFDLDIDPNILEYNFFETTIFKAPANLGYARAGVDPPPCDA
GTDQDSRWRLQYDVYQYFLPENDLTEEMLLKHLQRMVSVVPQVKASALKVVTLTAND
KTSVFSFSLPGQGVYINVVDPLNTSAAYIPAHTYACSFAGEGSCASLGRVSSKVFV
TLFALLGGFFICFFHGRFWKTELFIFGIIMGFFYILITRLPIKYDVNLLITAVTGSVGMFMF
LVAVVWVRFGILSICMLCVGLVLGLFISSVTFTPLGNLKFHDDGFWVTFSCIALIPVV

FMGCLRILNILTCGVIGSYSVVLAIIDSYWSTSLSYITLNVLKRALNKDFHRAFTNVPFQT
NDFIILAVWGMLAVSGITLQIRREGRPFPPHPYKLWKQERERRVTNILDPSYHIPPLRE
RLYGRLTQIKGLFQKEQPAGERTPLLL

[0243] Modifications may be made to the foregoing without departing from the basic aspects of the invention. Although the invention has been described in substantial detail with reference to one or more specific embodiments, those of skill in the art will recognize that changes may be made to the embodiments specifically disclosed in this application, yet these modifications and improvements are within the scope and spirit of the invention, as set forth in the aspects which follow. All publications or patent documents cited in this specification are incorporated herein by reference as if each such publication or document was specifically and individually indicated to be incorporated herein by reference.

[0244] Citation of the above publications or documents is not intended as an admission that any of the foregoing is pertinent prior art, nor does it constitute any admission as to the contents or date of these publications or documents. U.S. patents and other publications referenced herein are hereby incorporated by reference.

What is claimed is:

1. A method for identifying a subject at risk of osteoarthritis, which comprises detecting the presence or absence of one or more polymorphic variations associated with osteoarthritis in a nucleic acid sample from a subject, wherein the one or more polymorphic variations are detected in a nucleotide sequence selected from the group consisting of:

(a) a nucleotide sequence in SEQ ID NO: 1-2;

(b) a nucleotide sequence which encodes a polypeptide encoded by a nucleotide sequence in SEQ ID NO: 1-2;

(c) a nucleotide sequence which encodes a polypeptide that is 90% or more identical to the amino acid sequence encoded by a nucleotide sequence in SEQ ID NO: 1-2;

(d) a fragment of a nucleotide sequence of (a), (b), or (c);

whereby the presence of the polymorphic variation is indicative of the subject being at risk of osteoarthritis.

2. The method of claim 1, which further comprises obtaining the nucleic acid sample from the subject.

3. The method of claim 1, wherein the one or more polymorphic variations are detected within a region spanning chromosome positions 27052000 to 27066000 in human genomic DNA.

4. The method of claim 1, wherein the one or more polymorphic variations are detected at one or more positions selected from the group consisting of 230, 231, 5330, 6334, 11372, 11456, 11501, 13393, 16666, 17596, 19710, 19800, 20297, 20967, 32514, 33159, 37600, 41259, 41329, 50060, 53292, 53393, 56417, 56435, 58847, 59595, 59661, 60355, 60407, 62357, 68230, 68516, 69055, 72603, 73928, 85897 and 91554.

5. The method of claim 1, wherein the one or more polymorphic variations are detected at one or more positions in SEQ ID NO: 1 selected from the group consisting of 56435, 59595, 53292, 33159 and 41329.

6. The method of claim 1, wherein the one or more polymorphic variations are detected at one or more positions in linkage disequilibrium with one or more positions in claim 3, 4 or 5.

7. The method of claim 1, wherein detecting the presence or absence of the one or more polymorphic variations comprises:

hybridizing an oligonucleotide to the nucleic acid sample, wherein the oligonucleotide is complementary to a nucleotide sequence in the nucleic acid and hybridizes to a region adjacent to the polymorphic variation;

extending the oligonucleotide in the presence of one or more nucleotides, yielding extension products; and

detecting the presence or absence of a polymorphic variation in the extension products.

8. The method of claim 1, wherein the subject is a human.

9. The method of claim 8, wherein the subject is a human female.

10. The method of claim 8, wherein the subject is a human male.

11. A method for identifying a polymorphic variation associated with osteoarthritis proximal to an incident polymorphic variation associated with osteoarthritis, which comprises:

identifying a polymorphic variation proximal to the incident polymorphic variation associated with osteoarthritis, wherein the polymorphic variation is detected in a nucleotide sequence selected from the group consisting of:

(a) a nucleotide sequence in SEQ ID NO: 1-2;

(b) a nucleotide sequence which encodes a polypeptide encoded by a nucleotide sequence in SEQ ID NO: 1-2;

(c) a nucleotide sequence which encodes a polypeptide that is 90% or more identical to the amino acid sequence encoded by a nucleotide sequence in SEQ ID NO: 1-2;

(d) a fragment of a nucleotide sequence of (a), (b), or (c) comprising a polymorphic variation;

determining the presence or absence of an association of the proximal polymorphic variant with osteoarthritis.

12. The method of claim 11, wherein the incident polymorphic variation is at one or more positions in claim 3, 4 or 5.

13. The method of claim 11, wherein the proximal polymorphic variation is within a region between about 5 kb 5' of the incident polymorphic variation and about 5 kb 3' of the incident polymorphic variation.

14. The method of claim 11, which further comprises determining whether the proximal polymorphic variation is in linkage disequilibrium with the incident polymorphic variation.

15. The method of claim 11, which further comprises identifying a second polymorphic variation proximal to the identified proximal polymorphic variation associated with osteoarthritis and determining if the second proximal polymorphic variation is associated with osteoarthritis.

16. The method of claim 15, wherein the second proximal polymorphic variant is within a region between about 5 kb 5' of the incident polymorphic variation and about 5 kb 3' of the proximal polymorphic variation associated with osteoarthritis.

17. An isolated nucleic acid comprising a nucleotide sequence selected from the group consisting of:

- (a) a nucleotide sequence in SEQ ID NO: 1-2;
 - (b) a nucleotide sequence which encodes a polypeptide encoded by a nucleotide sequence in SEQ ID NO: 1-2;
 - (c) a nucleotide sequence which encodes a polypeptide that is 90% or more identical to the amino acid sequence encoded by a nucleotide sequence in SEQ ID NO: 1-2;
 - (d) a fragment of a nucleotide sequence of (a), (b), or (c) comprising a polymorphic variation; and
 - (e) a nucleotide sequence complementary to the nucleotide sequences of (a), (b), (c), or (d);
- wherein the nucleotide sequence comprises a polymorphic variation associated with osteoarthritis selected from the group consisting of a thymine thymine repeat at position 56435, a thymine at position 59595, a cytosine at position 53292, a guanine at position 33159 and a thymine at position 41329.

18. An oligonucleotide comprising a nucleotide sequence complementary to a portion of the nucleotide sequence of (a), (b), (c), or (d) in claim 17, wherein the 3' end of the oligonucleotide is adjacent to a polymorphic variation associated with osteoarthritis.

19. A microarray comprising an isolated nucleic acid of claim 17 linked to a solid support.

20. An isolated polypeptide encoded by the isolated nucleic acid sequence of claim 17.

21. A method for identifying a candidate therapeutic for treating osteoarthritis, which comprises:

(a) introducing a test molecule to a system which comprises a nucleic acid comprising a nucleotide sequence selected from the group consisting of:

(i) a nucleotide sequence in SEQ ID NO: 1-2;

(ii) a nucleotide sequence which encodes a polypeptide encoded by a nucleotide sequence in SEQ ID NO: 1-2;

(iii) a nucleotide sequence which encodes a polypeptide that is 90% or more identical to the amino acid sequence encoded by a nucleotide sequence in SEQ ID NO: 1-2;

(iv) a fragment of a nucleotide sequence of (a), (b), or (c); or

introducing a test molecule to a system which comprises a protein encoded by a nucleotide sequence of (i), (ii), (iii), or (iv); and

(b) determining the presence or absence of an interaction between the test molecule and the nucleic acid or protein,

whereby the presence of an interaction between the test molecule and the nucleic acid or protein identifies the test molecule as a candidate therapeutic for treating osteoarthritis.

22. The method of claim 21, wherein the system is an animal.

23. The method of claim 21, wherein the system is a cell.

24. The method of claim 21, wherein the nucleotide sequence comprises one or more polymorphic variations associated with osteoarthritis.

25. The method of claim 24, wherein the one or more polymorphic variations associated with osteoarthritis are at one or more positions in claim 3, 4 or 5.

26. A method for treating osteoarthritis in a subject, which comprises contacting one or more cells of a subject in need thereof with a nucleic acid, wherein the nucleic acid comprises a nucleotide sequence selected from the group consisting of:

(a) a nucleotide sequence in SEQ ID NO: 1-2;

(b) a nucleotide sequence which encodes a polypeptide encoded by a nucleotide sequence in SEQ ID NO: 1-2;

(c) a nucleotide sequence which encodes a polypeptide that is 90% or more identical to the amino acid sequence encoded by a nucleotide sequence in SEQ ID NO: 1-2;

(d) a fragment of a nucleotide sequence of (a), (b), or (c); and

(e) a nucleotide sequence complementary to the nucleotide sequences of (a), (b), (c), or (d);

whereby contacting the one or more cells of the subject with the nucleic acid treats the osteoarthritis in the subject.

27. The method of claim 26, wherein the nucleic acid is RNA or PNA.

28. The method of claim 27, wherein the nucleic acid is duplex RNA.

29. A method for treating osteoarthritis in a subject, which comprises contacting one or more cells of a subject in need thereof with a protein, wherein the protein is encoded by a nucleotide sequence which comprises a polynucleotide sequence selected from the group consisting of:

(a) a nucleotide sequence in SEQ ID NO: 1-2;

(b) a nucleotide sequence which encodes a polypeptide encoded by a nucleotide sequence in SEQ ID NO: 1-2;

(c) a nucleotide sequence which encodes a polypeptide that is 90% or more identical to the amino acid sequence encoded by a nucleotide sequence in SEQ ID NO: 1-2;

(d) a fragment of a nucleotide sequence of (a), (b), or (c);

whereby contacting the one or more cells of the subject with the protein treats the osteoarthritis in the subject.

30. A method for treating osteoarthritis in a subject, which comprises:

detecting the presence or absence of one or more polymorphic variations associated with osteoarthritis in a nucleic acid sample from a subject, wherein the one or more polymorphic variation are detected in a nucleotide sequence selected from the group consisting of:

(a) a nucleotide sequence in SEQ ID NO: 1-2;

(b) a nucleotide sequence which encodes a polypeptide encoded by a nucleotide sequence in SEQ ID NO: 1-2;

(c) a nucleotide sequence which encodes a polypeptide that is 90% or more identical to the amino acid sequence encoded by a nucleotide sequence in SEQ ID NO: 1-2;

(d) a fragment of a nucleotide sequence of (a), (b), or (c) comprising a polymorphic variation; and

administering an osteoarthritis treatment to a subject in need thereof based upon the presence or absence of the one or more polymorphic variations in the nucleic acid sample.

31. The method of claim 30, wherein the one or more polymorphic variations are detected at one or more positions in claim 3, 4 or 5.

32. The method of claim 30, wherein the treatment is selected from the group consisting of administering a corticosteroid, a corticosteroid, a nonsteroidal anti-inflammatory drug (NSAID), a cyclooxygenase-2 (COX-2) inhibitor, an antibody, a glucocorticoid, hyaluronic acid, chondroitin sulfate, glucosamine or acetaminophen; prescribing a heat/cold regimen or a joint protection regimen; performing joint surgery; prescribing a weight control regimen; and combinations of the foregoing.

33. A method for detecting or preventing osteoarthritis in a subject, which comprises:
detecting the presence or absence of one or more polymorphic variations associated with osteoarthritis in a nucleic acid sample from a subject, wherein the polymorphic variation is detected in a nucleotide sequence selected from the group consisting of:

- (a) a nucleotide sequence in SEQ ID NO: 1-2;
- (b) a nucleotide sequence which encodes a polypeptide encoded by a nucleotide sequence in SEQ ID NO: 1-2;
- (c) a nucleotide sequence which encodes a polypeptide that is 90% or more identical to the amino acid sequence encoded by a nucleotide sequence in SEQ ID NO: 1-2;
- (d) a fragment of a nucleotide sequence of (a), (b), or (c) comprising a polymorphic variation; and

administering an osteoarthritis prevention or detection procedure to a subject in need thereof based upon the presence or absence of the one or more polymorphic variations in the nucleic acid sample.

34. The method of claim 33, wherein the one or more polymorphic variations are detected at one or more positions in claim 3, 4 or 5.

35. The method of claim 33, wherein the osteoarthritis prevention is selected from the group consisting of administering a corticosteroid, a nonsteroidal anti-inflammatory drug (NSAID), a cyclooxygenase-2 (COX-2) inhibitor, an antibody, a glucocorticoid, hyaluronic acid, chondroitin sulfate, glucosamine or acetaminophen; prescribing a heat/cold regimen or a joint protection regimen; performing joint surgery; prescribing a weight control regimen; and combinations of the foregoing.

36. A method of targeting information for preventing or treating osteoarthritis to a subject in need thereof, which comprises:

detecting the presence or absence of one or more polymorphic variations associated with osteoarthritis in a nucleic acid sample from a subject, wherein the polymorphic variation is detected in a nucleotide sequence selected from the group consisting of:

(a) a nucleotide sequence in SEQ ID NO: 1-2;

(b) a nucleotide sequence which encodes a polypeptide encoded by a nucleotide sequence in SEQ ID NO: 1-2;

(c) a nucleotide sequence which encodes a polypeptide that is 90% or more identical to the amino acid sequence encoded by a nucleotide sequence in SEQ ID NO: 1-2;

(d) a fragment of a nucleotide sequence of (a), (b), or (c) comprising a polymorphic variation; and

directing information for preventing or treating osteoarthritis to a subject in need thereof based upon the presence or absence of the one or more polymorphic variations in the nucleic acid sample.

37. The method of claim 36, wherein the one or more polymorphic variations are detected at one or more positions in claim 3, 4 or 5.

38. A composition comprising a cell from a subject having osteoarthritis or at risk of osteoarthritis and an antibody that specifically binds to a protein, polypeptide or peptide encoded by a nucleotide sequence identical to or 90% or more identical to a nucleotide sequence in SEQ ID NO: 1-2.

39. A composition comprising a cell from a subject having osteoarthritis or at risk of osteoarthritis and a RNA, DNA, PNA or ribozyme molecule comprising a nucleotide sequence identical to or 90% or more identical to a portion of a nucleotide sequence in SEQ ID NO: 1-2.

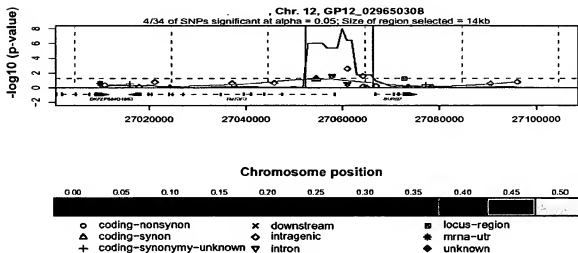
40. The composition of claim 39, wherein the RNA molecule is a short inhibitory RNA molecule.

Abstract of the Disclosure

Provided herein are methods for identifying a risk of osteoarthritis in a subject, reagents and kits for carrying out the methods, methods for identifying candidate therapeutics for treating osteoarthritis, and therapeutic and preventative methods applicable to osteoarthritis. These embodiments are based upon an analysis of polymorphic variations in nucleotide sequences within the human genome.

FIGURE 1

TM7SF3 – DISCOVERY P-VALUES (female only)



Application Data Sheet

Application Information

Application Type::	Provisional
Subject Matter::	Utility
Suggested Group Art Unit::	Not Yet Assigned
CD-ROM or CD-R?::	None
Sequence submission?::	None
Computer Readable Form (CRF)?::	No
Title::	METHODS FOR IDENTIFYING RISK OF OSTEOARTHRITIS AND TREATMENTS THEREOF
Attorney Docket Number::	524593009100
Request for Early Publication?::	No
Request for Non-Publication?::	No
Total Drawing Sheets?::	1
Small Entity?::	Yes
Petition included?::	No
Secrecy Order in Parent Appl.?::	No

Applicant Information

Applicant Authority Type::	Inventor
Primary Citizenship Country::	US
Status::	Full Capacity
Given Name::	Steven
Family Name::	MAH
City of Residence::	San Diego
State or Province of Residence::	CA
Country of Residence::	US
Street of mailing address::	12820 Via Nieve #74
City of mailing address::	San Diego
State or Province of mailing address::	CA

Postal or Zip Code of mailing address:: 92130

Applicant Authority Type:: Inventor
Primary Citizenship Country:: Germany
Status:: Full Capacity
Given Name:: Andreas
Family Name:: BRAUN
City of Residence:: San Diego
State or Province of Residence:: CA
Country of Residence:: US
Street of mailing address:: 3935 Lago Di Grata Circle
City of mailing address:: San Diego
State or Province of mailing address:: CA
Postal or Zip Code of mailing address:: 92130

Applicant Authority Type:: Inventor
Primary Citizenship Country:: Germany
Status:: Full Capacity
Given Name:: Stefan
Middle Name:: M.
Family Name:: KAMMERER
City of Residence:: San Diego
State or Province of Residence:: CA
Country of Residence:: US
Street of mailing address:: 3825 Elijah Court, Unit 334
City of mailing address:: San Diego
State or Province of mailing address:: CA
Postal or Zip Code of mailing address:: 92130

Applicant Authority Type:: Inventor
Primary Citizenship Country:: US
Status:: Full Capacity

Given Name:: Matthew
Middle Name:: Roberts
Family Name:: NELSON
City of Residence:: San Marcos
State or Province of Residence:: CA
Country of Residence:: US
Street of mailing address:: 1250 Calle Prospero
City of mailing address:: San Marcos
State or Province of mailing address:: CA
Postal or Zip Code of mailing address:: 92069

Applicant Authority Type:: Inventor
Primary Citizenship Country:: Sweden
Status:: Full Capacity
Given Name:: Rikard
Middle Name:: Henry
Family Name:: RENELAND
City of Residence:: San Diego
State or Province of Residence:: CA
Country of Residence:: US
Street of mailing address:: 7555 Charmant Drive, #1114
City of mailing address:: San Diego
State or Province of mailing address:: CA
Postal or Zip Code of mailing address:: 92122

Applicant Authority Type:: Inventor
Primary Citizenship Country:: United Kingdom
Status:: Full Capacity
Given Name:: Maria
Middle Name:: L.
Family Name:: LANGDOWN
City of Residence:: San Diego

State or Province of Residence:: CA
Country of Residence:: US
Street of mailing address:: 3701 Yosemite Street
City of mailing address:: San Diego
State or Province of mailing address:: CA
Postal or Zip Code of mailing address:: 92109

Correspondence Information

Correspondence Customer Number:: 25225

Representative Information

Representative Customer Number:: 25225